OM protein - protein search, using sw model

```
                              GenCore version 5.1.6
                              Copyright (c) 1993 - 2006 Compugen Ltd.

Run on:              January 11, 2006, 01:53:16 ; Search time 135 Seconds
                                                  (without alignments)
                                                  732.298 Million cell updates/sec

Title:               US-10-006-922A-12
Perfect score:       1214
Sequence:            1 MRSSKNVIKEFMRFKVRMEG..........EDYTIVEQYERTEGRHHLFL 225
Scoring table:       BLOSUM62
                     Gapop 10.0 , Gapext 0.5
Searched:            2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters:    2443163

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :     A_Geneseq_21:*
               1: geneseqp1980s:*
               2: geneseqp1990s:*
               3: geneseqp2000s:*
               4: geneseqp2001s:*
               5: geneseqp2002s:*
               6: geneseqp2003s:*
               7: geneseqp2003bs:*
               8: geneseqp2004s:*
               9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	1214	100.0	225	3	AAY99836	Aay99836 Discosoma
2	1214	100.0	225	3	AAB01622	Aab01622 Discosoma
3	1214	100.0	225	4	AAG65509	Aag65509 Anthozoan
4	1214	100.0	225	5	ABB08834	Abb08834 Yeast opt
5	1214	100.0	225	5	AAE28833	Aae28833 Discosoma
6	1214	100.0	225	5	AAB17540	Aab17540 Discosoma
7	1214	100.0	225	5	AAO18270	Aao18270 Discosoma
8	1214	100.0	225	6	ADY51716	Ady51716 Discosoma
9	1214	100.0	225	6	AAE34962	Aae34962 Discosoma
10	1214	100.0	225	7	ADC24126	Adc24126 Discosoma
11	1214	100.0	225	7	ABW00918	Abw00918 Discosoma
12	1214	100.0	225	7	ADF70403	Adf70403 Discosoma
13	1214	100.0	225	7	ADH34489	Adh34489 Discosoma
14	1214	100.0	225	7	ADL46203	Adl46203 Discosoma
15	1214	100.0	225	7	ADN33979	Adn33979 Wild-type
16	1214	100.0	225	8	ADI36421	Adi36421 Discosoma
17	1214	100.0	225	7	ADM97769	Adm97769 D sp red
18	1214	100.0	225	8	ADQ28780	Adq28780 Jellyfish
19	1214	100.0	225	9	ADX26534	Adx26534 Discosoma
20	1214	100.0	487	5	ABB08821	Abb08821 Autofluor
21	1214	100.0	487	5	ABB08822	Abb08822 Autofluor
22	1214	100.0	506	5	ABB08823	Abb08823 Autofluor
23	1211	99.8	547	5	AAE28920	Aae28920 Discosoma
24	1211	99.8	225	7	ABW00937	Abw00937 Discosoma
25	1211	99.8	225	7	ABW00929	Abw00929 Discosoma
26	1211	99.8	225	7	ABW00938	Abw00938 Discosoma
27	1211	99.8	225	7	ADH34498	Adh34498 Discosoma
28	1211	99.8	225	7	ADH34499	Adh34499 Discosoma
29	1211	99.8	225	5	AAE28919	Aae28919 Discosoma
30	1210	99.7	225	5	ABW00930	Abw00930 Discosoma
31	1210	99.7	226	4	AAG65510	Aag65510 Anthozoan
32	1210	99.7	242	7	ADZ24109	Adz24109 Discosoma
33	1210	99.7	545	7	ADI18132	Adi18132 RFP:PS [NI
34	1210	99.7	548	7	ADI18156	Adi18156 RFP:PS [HI
35	1209	99.6	225	7	AAE28922	Aae28922 Discosoma
36	1209	99.6	225	7	ABW00936	Abw00936 Discosoma
37	1209	99.6	225	7	ABW00931	Abw00931 Discosoma
38	1209	99.6	226	5	ABB08835	Abb08835 Yeast opt
39	1208	99.5	225	7	ABW00932	Abw00932 Discosoma
40	1208	99.5	225	7	ABW00935	Abw00935 Discosoma
41	1208	99.5	225	7	ABW00939	Abw00939 Discosoma
42	1208	99.5	225	7	ADH34500	Adh34500 Discosoma
43	1207	99.4	225	5	AAE28921	Aae28921 Discosoma
44	1207	99.4	225	5	AAE17541	Aae17541 Discosoma
45	1207	99.4	225	5	ADY51734	Ady51734 Discosoma

ALIGNMENTS

```
RESULT 1
AAY99836
ID   AAY99836 standard; protein; 225 AA.
XX
AC   AAY99836;
XX
DT   12-SEP-2003  (revised)
DT   19-SEP-2000  (first entry)
XX
DE   Discosoma sp. "red" novel fluorescent protein drFP583.
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling.
XX
OS   Discosoma sp; "red".
XX
FH   Key             Location/Qualifiers
FH
FT   Misc-difference 122
FT                   /note= "encoded by TC"
FT   Misc-difference 127
FT                   /note= "encoded by GTTG"
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US029473.
PR
PR   11-DEC-1998;   98US-00210330.
PR   14-OCT-1999;   99US-00418529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA,  Fradkov AF,  Labas YA,  Matz MV,  Green G,  Chen Y;
PI   Ding L;
XX
PT   WPI; 2000-423381/36.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 20; Page 74-75; 86pp; English.
XX
CC   The present sequence is a novel fluorescent protein (nFP) encoded by the
CC   full-length cDNA drFP583. drFP583 was isolated from Discosoma sp. "red",
CC   a non-bioluminescent species of the Class Anthozoa. Fluorescent proteins
CC   can be used in fluorescent labeling, a useful tool for marking a protein,
```

```
CC   cell or organism of interest. Unlike other markers used in protein
CC   labeling, such as beta-galactosidase and luciferase, fluorescence proteins
CC   do not require an exogenous cofactor or substrate. Methods involving
CC   fluorescent proteins are also less laborious and less difficult to
CC   control than the traditional methods of fluorescent labeling, where a
CC   protein of interest is purified and then covalently conjugated to a
CC   fluorophore derivative. Novel fluorescent proteins isolated from species
CC   of the Class Anthozoa can be used as markers for gene expression and
CC   protein localization studies, and in fluorescence resonance energy
CC   transfer (FRET) reactions. They have improved properties and better
CC   suitability for larger excitations compared to prior art fluorescent
CC   proteins such as green fluorescent protein. (Updated on 12-SEP-2003 to
CC   standardise OS field)
XX
SQ   Sequence 225 AA;

Query Match      100.0%;  Score 1214;  DB 3;  Length 225;
Best Local Similarity    100.0%;  Pred. No. 1e-127;
Matches 225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 2
ID   AAB01622
AC   AAB01622 standard; protein; 225 AA.
XX
AC   AAB01622;
XX
DT   12-DEC-2000  (first entry)
XX
DE   Discosoma sp. red fluorescent protein drFP583.
XX
KW   Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;
KW   zFP506; zFP538; drFP583; dsFP483; asFP600; dgFP512; dmFP592.
XX
OS   Discosoma sp.
XX
PN   WO20003452G-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;  99WO-US029405.
XX
PR   11-DEC-1998;  98US-00210330.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
XX
DR   WPI; 2000-423451/36.
XX
PT   Novel method for identifying a DNA sequence encoding fluorescent proteins
PT   from non-bioluminescent Anthozoa which are useful for fluorescent
PT   labeling and as markers.
XX
PS   Claim 3; Page 68-69; 73pp; English.
XX
CC   The present sequence is Discosoma sp. red fluorescent protein drFP583. It
```

```
CC   was isolated using the Aequoria victoria green fluorescent protein (GPP)
CC   sequence, which was used to design PCR primers which might isolate other
CC   fluorescent proteins from a number of species of Anthozoa. These were
CC   Anemonia majano, Clavularia sp., Zoanthus sp., Discosoma sp. and Anemonia
CC   sulcata. The cDNA obtained was then screened in the search for sequences
CC   encoding fluorescent proteins. The other proteins found in this manner
CC   were cFP484, zFP506, zFP538, amFP486, dsFP483, asFP600, dgFP512 and
CC   dmFP592. These proteins can be used as fluorescent labels (for gene
CC   expression and protein localisation studies and in fluorescence resonance
CC   energy transfer (FRET) studies) in place of fluorophore derivatives and
CC   luciferases, as these involve laborious processes and the latter require
CC   cofactors. They can also be used in place of GFP, which is too stable to
CC   be useful when studying short-term or repetitive events
XX
SQ   Sequence 225 AA;

Query Match      100.0%;  Score 1214;  DB 3;  Length 225;
Best Local Similarity    100.0%;  Pred. No. 1e-127;
Matches 225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 3
ID   AAG65509
AC   AAG65509 standard; protein; 225 AA.
XX
AC   AAG65509;
XX
DT   30-NOV-2001  (first entry)
XX
DE   Anthozoan red fluorescent protein sequence.
XX
KW   Fluorescent protein; Anthozoan; fluorescence; marker; FRET; red.
XX
OS   Anthozoa.
XX
PN   WO200162919-A1.
XX
PD   30-AUG-2001.
XX
PF   13-FEB-2001;  2001WO-US004625.
XX
PR   23-FEB-2000;  2000US-0184732P.
XX
PA   (AURO-) AURORA BIOSCIENCES CORP.
XX
PI   Nelson D, Zamaira E, Tsien R;
XX
DR   WPI; 2001-557704/62.
XX
PT   Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
PT   functional red fluorescent proteins, and the encoding nucleic acids, with
PT   key mutations for improving the proteins function.
XX
PS   Disclosure; Page 85; 90pp; English.
XX
CC   The invention provides a nucleic acid encoding functional red fluorescent
CC   protein (II) that differs from the sequence of an Anthozoan red
```

```
CC   fluorescent protein by at least one amino acid substitution, and with
CC   different fluorescent properties. The red fluorescent protein of the
CC   invention can be expressed by standard recombinant methodology. (II) are
CC   used a fluorescent markers and FRET partners. It is used for identifying
CC   protein-protein interactions. (II) is also suitable for multiplexed
CC   fluorescent analysis and FRET-based applications using existing Aequorea
CC   fluorescent proteins. (II) has improved brightness, reduced spectral
CC   cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC   The key mutations in the encoding nucleic acids provide improved folding,
CC   brightness, and create (II) with sharper, more defined excitation and
CC   emission peaks when expressed in anthozoan fluorescent protein
CC   represents an anthozoan fluorescent protein
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 4;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 4
ID   ABB08834 standard; protein; 225 AA.
XX
AC   ABB08834;
XX
DT   29-MAY-2002  (first entry)
XX
DE   Yeast optimised RFP SEQ ID NO 17.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology.
XX
OS   Anthozoa.
XX
PN   DE20001395-U1.
XX
PD   15-MAR-2001.
XX
PF   27-JAN-2000; 2000DE-02001395.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
DR   WPI; 2002-228394/29.
DR   N-PSDB; ABA95905, ABA95921, ABA95922.
XX
PT   New DNA encoding red fluorescent protein, useful as marker in
PT   biotechnology, has sequence optimised for expression in eukaryotes,
PT   especially yeast or plants.
XX
PS   Disclosure; Page 13-14; 19pp; German.
XX
CC   The invention relates to DNA (I) containing either sequence ABA95905 or
CC   sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
```

```
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC   plants, especially dicotyledonous plants including Nicotiana tabacum or
CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern
CC   biotechnology. (I) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio. The present
CC   sequence is that of the yeast optimised RFP
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 5;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 5
ID   AAE28833 standard; protein; 225 AA.
XX
AC   AAE28833;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) wild-type protein.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   bioosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6.
XX
OS   Discosoma sp.
XX
PN   WO20020484559-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   N-PSDB; AAD46278.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
```

Disclosure; Page 70-71; 80pp; English.

XX

CC The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays, as colouring agents, as markers in
CC recombinant DNA applications, as sunscreens or filters, in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC sequence is Discosoma sp. drFP583 (NFP-6) wild-type protein of the
CC invention

XX

SQ Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No.1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 6
AAE17540
ID AAE17540 standard; protein; 225 AA.
XX
AC AAE17540;
XX
DT 22-APR-2002 (first entry)
XX
DE Discosoma sp. humanised wild-type Anthozoa protein drFP583.
XX
KW Fluorescent timer protein; protein movement; translocation; trafficking;
KW promoter activity; gene expression; transgenic plant; gene modification;
KW protein age; anthozoa protein; drFP583.
XX
OS Discosoma sp.
XX
PN WO20019637-A2.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US019097.
XX
PR 14-JUN-2000; 2000US-0211607P.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Fradkov AF, Terskikh A;
XX
PT New fluorescent timer proteins comprising an emission spectrum that
PT changes over time from a first wavelength to a second wavelength, useful
PT for monitoring intracellular protein movement, translocation, trafficking
PT or stability.
XX
PR WPI; 2002-154595/20.
DR N-PSDB; AAD28207.
XX

Example 1; Fig 1; 89pp; English.

XX
CC The invention relates to a fluorescent timer protein having an emission
CC spectrum that changes over time after synthesis from a first wavelength
CC to a second wavelength. The fluorescent timer proteins are useful in
CC monitoring the activity of a promoter, determining the age of a protein,
CC identifying an agent that modulates the activity of a promoter and in
CC enriching a population of cells comprising a fluorescent timer protein.
CC The fluorescent timer proteins are also useful for assessing gene
CC expression during development of a multicellular organism or during
CC cellular differentiation, in response to a drug or other inducer of
CC promoter activity, as a reporter to serve as a read-out of promoter
CC activity, monitoring intracellular protein movement or translocation,
CC protein trafficking, or protein stability, to investigate temporal
CC aspects of the activity of a regulatory element, for determining cell
CC fate during development and organ remodelling, in spatial and temporal
CC visualisation of newly synthesised proteins and accumulated proteins, and
CC in distinguishing between newly formed and pre-existing structures, e.g.
CC membrane junctions and extracellular matrix components. The fluorescent
CC timer proteins may further be used to investigations where photobleaching
CC techniques are employed, as detectable labels, as selectable markers, as
CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC assays, and as second messenger detectors. The nucleic acids can be used
CC to generate transgenic, non-human plants or animals or site-specific gene
CC modifications in cell lines. The present sequence is Discosoma sp.
CC humanised wild-type Anthozoa protein drFP583 used for generating
CC fluorescent proteins

XX

SQ Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No.1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 7
AAO18270
ID AAO18270 standard; protein; 225 AA.
XX
AC AAO18270;
XX
DT 26-SEP-2002 (first entry)
XX
DE Discosoma red fluorescent protein.
XX
KW Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
KW modified yeast strain; environmental pollution.
XX
OS Discosoma sp.
XX
PN DE10061872-A1.
XX
PD 20-JUN-2002.
XX
PF 12-DEC-2000; 2000DE-01061872.
XX
PR 12-DEC-2000; 2000DE-01061872.
XX

PA (LICH/) LICHTENBERG-PRATE H.
XX
PI Lichtenberg-Prate H;
XX
DR WPI; 2002-539633/58.
DR N-PSDB; AAL47952.
XX
PT Modified yeast strain, useful for detecting toxic compounds in
PT environment, contains integrated cassettes responsive to genotoxic and
PT cytotoxic compounds.
XX
PS Disclosure; Page 21-22; 34pp; German.
XX
CC The present invention relates to a modified yeast strain that contains,
CC integrated stably and functionally in its genome, a genotoxicity cassette
CC and a cytotoxicity cassette, each comprising a promoter and reporter
CC gene, both of which are different in the two cassettes. The modified
CC yeast strain is used to detect environmental pollution, especially
CC genotoxic and/or cytotoxic substances in complex environmental
CC contaminants, especially organic compounds, but also (non-)ionising
CC radiation and chemical carcinogens. Particular applications are in
CC monitoring (waste) water (e.g. as an early warning system), medical
CC toxicology screening and for industrial process control. The present
CC sequence is a marker protein suitable for use in the cassettes of the
CC present invention
XX
SQ Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 8
ID ADY51716
AC ADY51716;
XX
DT 05-MAY-2005 (first entry)
XX
DE Discosoma sp. wild type red fluorescent protein (RFP) Seq 12.
XX
KW fluorescence; mutagenesis; red fluorescent protein; protein interaction.
XX
OS Discosoma sp.
XX
PN WO200268605-A2.
XX
PD 06-SEP-2002.
XX
PF 26-FEB-2002; 2002WO-US006063.
PR 26-FEB-2001; 2001US-00794308.
PR 24-MAY-2001; 2001US-0086538.
XX
PA (REGC) UNIV CALIFORNIA.

XX
PI Tsien RY, Baird GS, Campbell RE, Zacharias DA;
XX
DR WPI; 2002-713372/77.
DR N-PSDB; ADY51715.
XX
PT New non-oligomerizing fluorescent protein containing at least one
PT mutation that reduces or eliminates the ability of the protein to
PT oligomerize, useful for making better and new assays for molecular
PT biology.
XX
SQ Claim 10; SEQ ID NO 12; 117pp; English.
XX
CC This invention relates to a novel non-oligomerizing fluorescent protein.
CC Specifically, it refers to the presence of at least one mutation in the
CC fluorescent protein that reduces or eliminates the ability of the protein
CC to oligomerize. The present invention describes fluorescent proteins
CC derived from naturally occurring green or red fluorescent proteins and
CC provides a fusion protein that comprises a non-oligomerizing fluorescent
CC protein linked to at least one protein of interest. As such, these fusion
CC proteins can be used in methods and compositions to determine the pH of a
CC sample, or whether the sample contains an enzyme, molecule or agent that
CC regulates the activity of an expression control sequence. Furthermore,
CC they may be used to identify a specific interaction of molecules, such
CC that they are useful for improving or developing new assays in the field
CC of molecular biology. This polypeptide sequence is the Discosoma sp. wild
CC type red fluorescent protein (RFP) of the invention.
XX
SQ Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 9
ID AAE34962
AC AAE34962;
XX
DT 28-MAY-2003 (first entry)
XX
DE Discosoma species red fluorescent protein (RFP).
XX
KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
KW kinase; red fluorescent protein; RFP.
XX
OS Discosoma sp.
XX
PN WO200295058-A2.
XX
PD 28-NOV-2002.
XX
PF 24-MAY-2002; 2002WO-US016955.
XX
PR 24-MAY-2001; 2001US-00865291.

XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Tsien RY, Ting AY, Zhang J;
XX
DR   WPI; 2003-148474/14.
DR   N-PSDB; AAD53432.
XX
PT   Novel chimeric phosphorylation indicators, useful for detecting
PT   kinase/phosphatase in samples, has donor molecule, phosphorylatable
PT   domain, phosphoaminoacid binding domain, and acceptor molecule, in
PT   operative linkage.
XX
PS   Disclosure; Col 65-66; 38pp; English.
XX
CC   The present invention relates to chimeric phosphorylation indicators
CC   comprising a phosphorylation polypeptide and a fluorescent protein or in
CC   operative linkage, a donor molecule, a phosphorylatable domain, a
CC   phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The
CC   phosphorylation indicators of the invention are useful for detecting
CC   kinases or phosphatases in a biological sample. They are also useful in
CC   high throughput analysis e.g. for detecting a kinase inhibitor or
CC   phosphatase inhibitor. The present sequence is Discosoma species red
CC   fluorescent protein (RFP) used in the invention
XX
SQ   Sequence 225 AA;
XX
Query Match   100.0%; Score 1214; DB 6; Length 225;
Best Local Similarity   100.0%; Pred. No. 1e-127;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
             |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 10
ADC24126 standard; protein; 225 AA.
ID   ADC24126;
AC   ADC24126;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Discosoma wild-type red fluorescent protein.
XX
KW   Discosoma red fluorescent protein variant; DsRed; AB interface; AC interface;
KW   fluorescent protein variant; transcription induction detection;
KW   fluorescence energy resonance transfer; FRET; protein kinase;
KW   protein phosphatase; ion indicator.
XX
OS   Discosoma.
XX
PN   US20030598035-A1.
XX
PD   27-MAR-2003.
XX
PF   10-APR-2002; 2002US-00121258.
PR   26-FEB-2001; 2001US-00794308.
PR   24-MAY-2001; 2001US-0086S538.

XX
PA   (TSIE/) TSIEN R Y.
PA   (CAMP/) CAMPBELL R E.
XX
PI   Tsien RY, Campbell RE;
XX
DR   WPI; 2003-743764/70.
DR   N-PSDB; ADC24127, ADC24134.
XX
PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.
XX
Claim 1; SEQ ID NO 1; 67pp; English.
XX
CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (SI) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   conaining a vector which comprises (I) operatively linked to an
CC   expression control sequence, and an unit to assay the variant fluorescent
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcription, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This is the amino acid sequence of
CC   Discosoma wild-type red fluorescent protein.
XX
SQ   Sequence 225 AA;
XX
Query Match   100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity   100.0%; Pred. No. 1e-127;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
             |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 11
ABW00918 standard; protein; 225 AA.
ID   ABW00918;
AC   ABW00918;

DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP.
XX
KW   Fluorescent protein; resonance energy transfer (RFP).
XX
PN   Discosoma sp.
XX
OS   Discosoma sp.
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
PR   26-FEB-2001; 2001US-00794308.
XX
PI   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
SQ   Sequence 225 AA;

Claim 10; Page 30-31; Opp; English.

Fluorescent proteins containing a mutation that reduces or eliminates its
ability to oligomerize which gives more reliable fluorescence resonance
energy transfer results and are useful to detect molecule interaction,
enzymes, or sample pH.

CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP)

DR   WPI; 2003-802418/75.
DR   N-PSDB; AAD61969.

Query Match   100.0%;   Score 1214;   DB 7;   Length 225;
Best Local Similarity   100.0%;   Pred. No. 1e-127;
Matches 225;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 12
ID   ADF70403
AC   ADF70403;
DT   12-FEB-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein; 225 AA.
XX

DE   Discosoma wild-type GFP variant protein SeqID26.
XX
KW   ligand; orphan receptor protein; fusion protein; fluorescent protein;
KW   cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
KW   GFPuv; Enhanced GFP; EGFP.
XX
OS   Discosoma sp.
XX
PN   WO200307l2272-A1.
XX
PD   28-AUG-2003.
XX
PF   21-FEB-2003; 2003WO-JP001901.
PR   22-FEB-2002; 2002JP-00045728.
PR   23-JUL-2002; 2002JP-00213949.
PR   11-OCT-2002; 2002JP-00298237.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX
SQ   Sequence 225 AA;

Disclosure; SEQ ID NO 26; 594pp; Japanese.

This invention relates to a novel method of identifying ligands to an
orphan receptor protein which comprises transforming cells with a
DNA encoding a fusion protein of the orphan receptor with a fluorescent
protein, so that the fusion protein is expressed in the cells (or cell
membranes isolated from them) and contacting the cells with the potential
ligand to be tested. A suitable fluorescent protein for incorporation in
the fusion protein is green fluorescent protein (GFP), for example GFP-1,
wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
identification of ligands binding to an orphan receptor protein.

DR   WPI; 2003-697654/66.
DR   N-PSDB; ADF70404.

Query Match   100.0%;   Score 1214;   DB 7;   Length 225;
Best Local Similarity   100.0%;   Pred. No. 1e-127;
Matches 225;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 13
ID   ADH34489
AC   ADH34489;
DT   11-MAR-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein DsRed (wild-type).
XX

```
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   fluorescence activated cell sorting; fluorescent timer;
KW   red fluorescent protein; DsRed.
XX
OS   Discosoma sp.
XX
FH
FH   Key             Location/Qualifiers
FT   Misc-difference 146
FT                   /note= "Key position for chromoprotein/fluorescent
FT                   protein activity. Corresponds to GFP residue 148"
FT   Misc-difference 161
FT                   /note= "Key position for chromoprotein/fluorescent
FT                   protein activity. Corresponds to GFP residue 165"
FT   Misc-difference 163
FT                   /note= "Key position for chromoprotein/fluorescent
FT                   protein activity. Corresponds to GFP residue 165"
FT   Misc-difference 197
FT                   /note= "Key position for chromoprotein/fluorescent
FT                   protein activity. Corresponds to GFP residue 167"
FT   Misc-difference 197
FT                   /note= "Key position for chromoprotein/fluorescent
FT                   protein activity. Corresponds to GFP residue 203"
XX
PN   WO200305783-A2.
XX
PD   17-JUL-2003.
PF   23-DEC-2002; 2002WO-US041418.
XX
PR   26-DEC-2001; 2001US-0343128P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Fig 1; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably an Anthozoan species. The
CC   Cnidarian species, and most preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions is, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants are useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as sunscreens and selective filters. Chromoprotein
CC   mutants having fluorescent activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g. transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g. green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence
CC   represents a wild-type red fluorescent protein, DsRed, from Discosoma sp.
CC   that was used as a parent sequence for the generation of mutant proteins
CC   in an example of the invention.
XX
SQ   Sequence 225 AA;

Query Match    100.0%;  Score 1214;  DB 7;  Length 225;
Best Local Similarity    100.0%;  Pred. No. 1e-127;
Matches 225;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 14
ADL46203
ID   ADL46203 standard; protein; 225 AA.
XX
AC   ADL46203;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Discosoma red fluorescent (DsRed) protein.
XX
KW   red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
PN   WO2003086446-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010879.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-0020208.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Tsien RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
DR   N-PSDB; ADL46204.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PS   Claim 1; SEQ ID NO 1; 166pp; English.
```

XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   protein.
XX
SQ   Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity   100.0%;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 15
ID   ADN33979
AC   ADN33979
XX
DT   01-JUL-2004 (first entry)
XX
PN   ADN33979 standard; protein; 225 AA.
XX
PD   03-JUL-2003.
XX
PF   18-DEC-2002; 2002WO-US040539.
XX
PR   19-DEC-2001; 2001US-0341723P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Bevis B, Glick B;
XX
PN   WO200305415B-A2.
XX
OS   Discosoma sp.
XX
KW   Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED.
XX
DE   Wild-type DsRED protein.
XX
DR   WPI; 2003-569236/53.
DR   N-PSDB; ADN33978.
XX
PS   Claim 8; SEQ ID NO 2; 65pp; English.
XX
PT   Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT   mutant of a Cnidarian chromo- or fluorescent protein or its mutant,
PT   useful for applications involving chromo- or fluorescent proteins.

The present invention relates to nucleic acid that encodes a rapidly
maturing chromo or fluorescent mutant of a Cnidarian chromo- or
fluorescent protein or its mutant. The protein is useful in applications
involving nucleic acid encoding a chromo- or fluorescent protein which is
useful for producing a chromo and/or fluorescent protein and is
growing the cell, whereby the protein is expressed, and isolating the
protein substantially free of other proteins. The protein is useful as
applications involving chromo- or fluorescent protein and is useful as
PCR primers, hybridization probes, etc. The expression cassettes are
useful for synthesizing related proteins. The chromoproteins are useful
as coloring agents which are capable of imparting color or pigment to a
particular composition of matter e.g. food compositions, pharmaceuticals,
cosmetics, living organisms, e.g. animals and plants. The chromoproteins
may also find use as labels in analyte detection assays, e.g. assays for
biological analyses of interest and as selectable markers in recombinant
DNA applications, e.g. the production of transgenic cells and organisms.
The fluorescent proteins find use in a variety of different applications,
e.g. in fluorescence resonance energy transfer (FRET) applications, as
biosensors in prokaryotic and eukaryotic cells, in applications involving
the automated screening of arrays of cells expressing fluorescent
reporting groups by using microscopic imaging and electronic analysis, as
second messenger detectors, and in fluorescence activated cell sorting
applications and as in vivo marker in animals. The fluorescent proteins
also find use in protease cleavage assays. The proteins can also be used
is assays to determine the phospholipid composition in biological
membranes and as a fluorescent timer. The present sequence represents the
wild-type DsRED.
XX
SQ   Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity   100.0%;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Search completed: January 11, 2006, 02:02:58
Job time : 136 secs

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

Copyright (c) 1993 - 2006 Compugen Ltd.
GenCore version 5.1.6

Run on: January 11, 2006, 01:55:51 ; Search time 16 Seconds

Title: US-10-006-922A-12
Perfect Score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..........EDYIVEQYERTEGRHHLFL 225
(without alignments)
Scoring table: 1353.048 Million cell updates/sec

Searched: 283416 seqs, 9621673 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Databases :
1: PIR 80:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%	Query Match Length	DB	ID	Description
1	211.5	17.4	238	1	JQ1514	green-fluorescent
2	93.5	7.6	26926	1	I38344	titin, cardiac mus
3	92	7.6	1873	2	T30944	surface protein pr
4	91	7.5	458	1	A42386	hsp 90-binding pro
5	89	7.3	1433	1	A36734	bacillopeptidase F
6	87.5	7.2	1116	2	B70476	hypothetical prote
7	87	7.2	568	2	T06489	probable peptidylp
8	87	7.2	1484	2	C97196	probable membrane
9	85.5	7.0	456	2	G69397	signal-transducing
10	85	7.0	271	2	F69442	hypothetical prote
11	85	7.0	679	2	A40351	adhesion-type prot
12	84.5	7.0	725	1	IJMSNG	neural cell adhesi
13	84	6.9	1616	2	T17884	S-layer protein
14	83.5	6.9	268	2	E90276	conserved hypothet
15	83.5	6.9	340	2	E69544	hypothetical beta
16	83.5	6.9	374	2	T06245	hypothetical prote
17	83.5	6.9	374	2	T06244	gibberellin 3 beta
18	83	6.8	373	2	T50605	gibberellin 3 beta
19	83	6.8	15281	1	S41309	hypothetical prote
20	82.5	6.8	559	1	S55383	cyclosporin synthe
21	82.5	6.8	680	2	S17982	peptidylprolyl iso
22	82.5	6.8	680	2	T32919	Kallmann syndrome
23	82	6.8	931	2	B53125	hypothetical prote
24	81.5	6.7	292	2	C69106	restriction enzyme
25	81.5	6.7	296	2	T37989	glucose-1-phosphat
26	81.5	6.7	551	1	S72485	La 4.1 protein - h
27	81.5	6.7	629	2	C64180	peptidylprolyl iso
28	81.5	6.7	6805	2	S20901	hypothetical prote
29	81	6.7	346	2	S77025	titin - rabbit (fr
30	81	6.7	587	1	E69171	phosphoesterase-re
31	80.5	6.6	404	1	S03849	ribonucleoprotein
32	80.5	6.6	415	1	JC1494	ribonucleoprotein
33	80.5	6.6	862	2	F75116	hypothetical prote
34	80.5	6.6	1003	1	AJCHPR	phosphoribosylamin
35	80.5	6.6	1115	1	IJMSNL	neural cell adhesi
36	80	6.6	1115	2	A92250	phosphoribosyl (p
37	80	6.6	621	2	F98114	neural cell adhesi
38	80	6.6	690	2	A46372	choline binding pr
39	79.5	6.5	459	2	A46372	choline-binding pr
40	79.5	6.5	475	2	T44566	immunophilin FKBP5
41	79.5	6.5	487	2	T45982	conserved hypothet
42	79.5	6.5	648	2	B84139	hypothetical prote
43	79	6.5	853	1	IJBONC	ABC transporter (p
44	79	6.5	458	1	JN0873	neural cell adhesi
45	79	6.5	703	2	A64351	immunophilin p59 -
	79	6.5	1214	2	T47438	hypothetical prote
						disease resistance

ALIGNMENTS

RESULT 1
JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
green-fluorescent protein - Aequorea victoria
C;Species: Aequorea victoria
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 31-Dec-2004
C;Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331
R;Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A;Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A;Reference number: JQ1514; MUID:92175527; PMID:1347277
A;Accession: JS0692
A;Molecule type: mRNA
A;Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>
A;Cross-references: UNIPARC:UPI000173FD6; UNIPROT:Q17106; UNIPARC:UPI000173FD7; UNIPROT:Q17105; UNIPARC:UPI0001B7F4; PID:...
A;Accession: PQ0335
A;Molecule type: DNA
A;Residues: 1-107,'S',109-238 <PRA1>
A;Cross-references: UNIPROT:P42212; UNIPARC:UPI000002FB4; GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:...
A;Accession: JQ1514
A;Molecule type: protein
A;Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>
A;Cross-references: UNIPARC:UPI000002FB4; GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:...
A;Accession: S48693
A;Molecule type: preliminary
R;Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A;Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A;Reference number: S48693; MUID:94364470; PMID:8082767
A;Molecule type: mRNA
A;Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'R',...
A;Cross-references: UNIPARC:UPI0000007BC01; EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PII...
A;Experimental source: clone gfp1
A;Accession: S51331
A;Molecule type: mRNA
A;Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>
A;Cross-references: UNIPARC:UPI0000003582F; GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:...
R;Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A;Reference number: S51330
A;Accession: S51330
A;Molecule type: mRNA
A;Residues: 1-24,'Q',26-29,'R',31-83,'L',85-153,'G',155-156,'P',158-171,'K',173-208,'Q',...
A;Cross-references: UNIPARC:UPI000079E2D; EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PII...
A;Experimental source: clone gfp2
R;Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A65692; PDB:1GFL
A;Cross-references: UNIPARC:UPI000007BC01; EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PII...
R;Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A;Title: The molecular structure of green fluorescent protein.
A;Reference number: A65692; PDB:1GFL
A;Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
A;Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A',2-79,'R',81-9...

A;Reference number: A58953; MUID:92924543; PMID:9631087
A;Contents: annotation; X-ray crystallography, 1.9 angstroms
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emittin
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C;Genetics:
A;Gene: GFP
A;Introns: 69/3; 167/3
C;Keywords: chromoprotein; luminescence
C;Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match
Best Local Similarity   17.4%;   Score 211.5; DB 1; Length 238;
Matches   56;   Conservative   47;   Mismatches 102; Indels 13; Gaps   6;

Qy   16   VRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQYGSKVYVKH 75
                ::|:: ||||| |||||||||| ::|||| : || ::
Db   16   VELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT-GKLPVPWPTLVTTFSYGVQCFSRY 74

Qy   76   PADIP----DYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDG 133
               : |: ||::||:|||||| :::||||||| | | | ::|:|:||||::|::|||
Db   75   PDHMKQHDFFKSAMPEGYVQERTIFYKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG 134

Qy   134   PVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLK-----DGGHYLVEF--KSIYMAKKPV 187
               ||||||||||||||||||||||||||||||| : ||||||| : ::||||::|
Db   135   NILGHK-MEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPV 193

Qy   188   QLPGYYVVDSKLDIT----SHNEDYTIVEQYERTEGRHH 222
               |:||:|:||:|:| ::|::|:||::|:||:|:|||
Db   194   LLPDNHYLSTQSALSKDPNEKRDHMILLEFVTAAGITH 231

RESULT 2
I38344
titin, cardiac muscle [validated] - human
N;Alternate names: connectin
C;Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004
A;Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R;Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Reference number: A57430; MUID:96026330; PMID:7569978
A;Title: Titin: giant proteins in charge of muscle ultrastructure and elasticity.
A;Accession: I38344
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME
A;Molecule type: mRNA
A;Residues: 1-26926 <LAB1>
A;Residues: 1977-2014 <MUS>
A;Molecule type: mRNA
A;Cross-references: UNIPROT:Q10466; UNIPARC:UPI000016B0E9; EMBL:X83270; NID:g602579; PIDN:CAA58243.1; PI
A;Note: conformation and properties are reported for a synthetic peptide corresponding t
Biochemistry 34, 553-561, 1995
R;Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
R;Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380; PMID:1582406
A;Accession: S20898
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A;Cross-references: UNIPARC:UPI000016B0E7; EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID
A;Accession: S20897
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 15597-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A;Cross-references: UNIPARC:UPI000016B0E6; EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID
A;Residues: 16330-16382, 'S', ...
A;Molecule type: mRNA
A;Cross-references: UNIPARC:UPI000016B0E6; EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID

A;Accession: S20899
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 26729-26825 <KOL>
A;Cross-references: UNIPARC:UPI000017Z572; EMBL:X92412; NID:g1236761
R;Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A;Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiati
A;Reference number: S37393; MUID:94008990; PMID:8404852
A;Accession: S37393
A;Molecule type: mRNA
A;Residues: 26831-26926 <GAU>
A;Cross-references: UNIPARC:UPI000017Z573
R;Improta, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A;Reference number: A66736; PDB:1TIT
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R;Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A66201; PDB:1NCT
A;Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q32
C;Function:
A;Description: structural protein forming filaments in striated muscle
A;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prot
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
A;structural protein

Query Match
Best Local Similarity   7.7%;   Score 93.5; DB 1; Length 26926;
Matches   51;   Conservative   34;   Mismatches 84; Indels 77; Gaps 13;

Qy   15   KVRMEGTVNGHEFE----------IEGEGEGRPYEGHNTV----------------- 54
                | |||||||||||| || |||||||| |
Db 23493   RYRSTGLTEGLEYEHRVTAINARGSGKPSRPSKPIVAMDPIAPPGKPQONPRVTDTTRTSV 23552

Qy   55   PFAWDILSPQFQYGSKV--YV-------KH--------PADIPDYKKLSFPEGFKWE-RV 96
               || || :: | | || ||||||::||:||||||| ||
Db 23553   SLAWSV--PEDEGGSKVTGYLIEMQKDQHEWTKCNTTPTKIREYTLTHLPQGAEYRFRV 23610

Qy   97   MNFEDG---------GVVTVTQ--------DSSLQDGCFIYKVKFIGVNFPSDG-----PV 135
               ::|::| | || |||||||||||||||||||||| ||
Db 23611   LACNAGGPGEPAEVPGTVKVTEMLEYPDYELDERYQEGIFVRQGGVTRLTIPIKGKPFPI 23670

Qy   136   MQKKTMGWEASTERLYPRDGVLKGEIHKALKLK-----DGGHYLVEF-----KSIYMAK 184
               |:|| |||||||||||||||||||||||||| ||||||||| ||||||
Db 23671   CK------WTKEGQDISKRAMIATSETHTELVKEADRGDSGTYDLVLENKCGKKAVIKV 23725

Qy   185   KPVQLP 190
               ::|
Db 23726   RVIGSP 23731

RESULT 3

T30944
surface protein precursor - Enterococcus faecalis
C;Species: Enterococcus faecalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30944
R;Shankar, V.; Baghdayan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A;Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene enc
R;Reference number: 220943; MUID:99081742; PMID:9864215
A;Accession: T30944
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-1873 <SHA>
A;Cross-references: UNIPARC:Q924N7; UNIPROT:Q924N7; EMBL:AF034779; NID:g3873186;

```
Query Match          7.6%;  Score 92;  DB 2;  Length 1873;
Best Local Similarity  20.8%;  Pred. No. 17;
Matches  46;  Conservative  33;  Mismatches  56;  Indels  86;  Gaps  13;

Qy     30  EGEGEGERPYEGHNTVKLKVTKGGPLPFAWDILSPQPQYGS------------KVYVKHPA  77
Db    708  KGVGSEPITG-TIPIKIV---------------------YQDGSVGTTDLAVTVSKNIYENPGE  750

Qy     78  DIP-DYKKLSPPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDG-PV  135
Db    751  NIPAGYHKVTFTAG----------------EGTSIESGTTVFAVK-DGVSLPEDKGPLV  791

Qy    136  MQKKTMGWEASTERLYPRDGV---LKGE------IHKALKLKD---------GGHYLVEF--  177
Db    792  LKAK-------DGYDAKWPEEATQPIKADDTEFVSSATKLDDIIENPGAGYHKVTFTA  847

Qy    178  -----------KSIYMAKKPVQLPCGYYYVDSKLDITSHNEDYT  209
Db    848  GEGTSIESGTTVFAVKDGVSLP-----EDKLPVLKAKDGYT  883
```

RESULT 4
A42386
hsp 90-binding protein p59 - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A42386
R;Lebeau, M.C.; Massol, N.; Herrick, J.; Faber, L.E.; Renoir, J.M.; Radanyi, C.; Baulieu
J. Biol. Chem. 267, 4281-4284, 1992
A;Title: P59, an hsp 90-binding protein. Cloning and sequencing of its cDNA and preparat
R;Reference number: A42386; MUID:92165768; PMID:1537818
A;Accession: A42386
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-458 <LEB>
A;Cross-references: UNIPROT:P27124; UNIPARC:UPI000148548; GB:M84988; NID:g165594; PIDN:
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:83839)
A;Superfamily: human FK506-binding protein FKBP51; BKBP-type peptidylprolyl isomerase ho
F;50-97/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F;319-352/Domain: tetratricopeptide repeat homology <TT1>
F;353-386/Domain: tetratricopeptide repeat homology <TT2>

```
Query Match          7.5%;  Score 91;  DB 1;  Length 458;
Best Local Similarity  24.6%;  Pred. No. 3.1;
Matches  42;  Conservative  26;  Mismatches  49;  Indels  54;  Gaps  10;

Qy     57  AWDI----------LSPQPQYGSKVYVKHPADIPDYKKLSFPE---GFKWERVMNF  99
Db     89  AWDIAVATMKVGELCRITCKPEYAYGS---AGSPPKIPENATLVFEVELFEFKGEDLTDD  145

Qy    100  EDGGVVTVTQDSSLLQDGCFIYKVKFIGVNF--PSDGPVMQKKTMGWEASTERLYP----R  153
Db    146  EDGGI----------------IRRIRTRGEGYARPNDGAIVEVALEGY--YKDRLFPQRELR  189

Qy    154  DGVLKGE--------IHKAL-KLKDGGHYLVEFKSIY----MAKRPVQLPCGY  192
Db    190  FEVSGEGSLDLPCGLEKAIQRMEKGEHSILYLKPSYAFGNAGKEKPQIPPY  240
```

RESULT 5
A36734
bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A36734; A35131; A35750; B35750; S08223; JN0335; 139849; JU0084
A;Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning
A;Reference number: A35131; MUID:90170864; PMID:2106512
A;Accession: A35131
A;Molecule type: DNA
A;Residues: 1-365; 'S',367-682; EIMP','893,'Q',895-896 <SL2>
A;Cross-references: UNIPROT:P16397; UNIPARC:UPI000017C26; GB:M29035
A;Note: the authors translated the codon GAA for residue 545 as Leu
R;Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J. Biol. Chem. 265, 6845-6850, 1990
A;Title: Cloning, genetic organization, and characterization of a structural gene encodi
R;Reference number: A35750; MUID:90216713; PMID:2108961
A;Accession: A35750
A;Molecule type: DNA
A;Residues: 1-392; 'V',394-828; NIRTRLYSLKFCRSRHKSV; <WUA>
A;Cross-references: UNIPARC:UPI00016E7EA; GB:J05400; NID:g142607; PIDN:AAA83362.1; PID:
A;Note: this sequence has been corrected
A;Accession: B35750
A;Molecule type: DNA
A;Residues: 876-935,'CG' <WU2>
A;Cross-references: UNIPARC:UPI00000B1CF; GB:J05400; NID:g142607; PIDN:AAA83363.1; PID:
A;Note: this sequence has been corrected
R;Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A;Title: Nucleotide sequence of the sporulation gene spoIIGA from Bacillus subtilis.
A;Reference number: S08223; MUID:90174995; PMID:2106671
A;Accession: S08223
A;Molecule type: DNA
A;Residues: 1410-1433 <MAS>
A;Cross-references: UNIPARC:UPI00016B97A; EMBL:X17344; NID:g940165; PIDN:CAA35224.1; PID:
R;Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
A;Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric po
A;Reference number: JN0335; MUID:93005071; PMID:1368833
A;Accession: JN0335
A;Molecule type: protein
A;Residues: 195-218,'A' <KAT>
A;Cross-references: UNIPARC:UPI00000B2A24
A;Note: source of this material was Bacillus subtilis (natto)
R;Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A;Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli c
A;Reference number: 139846; MUID:89008108; PMID:3139638
A;Accession: 139849
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-211 <RES>
A;Cross-references: UNIPARC:UPI000016868; GB:M22630; NID:g142938; PIDN:AAA22458.1; PID:g
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
C.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, F.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel,

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, B.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69596
C;Genetics:
C;Experimental source: strain 168
A;Cross-references: UNIPARC:UPI00060385; GB:Z99111; GB:Z99112; GB:AL009126; NID:g26339
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1433 <KUN>
A;Superfamily: bacillopeptidase F; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-194/Domain: propeptide #status predicted <PRO>
F;195-1433/Product: bacillopeptidase F #status experimental <MAT>
F;195-1433/Domain: subtilisin homology <SBT>
F;227,274,452/Active site: Asp, His, Ser #status predicted
A;Map position: 135 (degrees)
C;Gene: bpr; bpf

Query Match   7.3%;   Score 89;   DB 1;   Length 1433;
Best Local Similarity   24.0%;   Pred. No. 21;
Matches   56;   Conservative   26;   Mismatches   71;   Indels   80;   Gaps   14;

Qy   17   RMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFA------------------ 57
           |::|| |:|| :|:| :| :|::|
Db   521   KAEGQVS-------VEGDDQEPVYQHEKV-TEAYEGGSLPLTLTAEDNVSVTSVKLSYKLD 574

Qy   58   ---WDILSPQFQYGSKVYVKHPADIPDYK---KLSPEGGFKWERVNNFEDGG------VVT 106
             :|:| ||| ::| : | ||| |: |
Db   575   QGEWTEITAKRISGDHLKGTYQAEIPDIKGTKLS---KLSPEGGFKWERVNNFEDGG---VVT 627

Qy   107   VTQDSSLQDGCFIYKVKF----------IGVN------FPSDGPVMQKKTMGWEASTERLY 151
            |||||| | | | |:| | : :|
Db   628   VTVKPSITAG-------YKQDFETAPGGWVASGTNNNWEWGVPSTGP-------NTAAGGEKVY 677

Qy   152   PRDGVLKGE----------IHKALKLKDGGHYLVIEFKSIYMAKKPVQLPGYYYV 195
            | | :||| |: ||| :| ||
Db   678   GTN---LTGNYANSANMNLVMPPIKAPDSGSLFLQFKSWHNLEDDFDY-GYVFV 727

Db   382   VVFQDFNAYLSGBGK 396
Qy   210   IVEQ-----YERTEGR 220

RESULT 6
B70476
hypothetical protein aq_2054 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: B70476
A;Molecule type: DNA
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70476
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1116 <AQF>
A;Cross-references: UNIPROT:067838; UNIPARC:UPI000056 7A2; GB:AE000770; NID:g2984274; PI
A;Experimental source: strain VF5
C;Genetics:
C;Gene: aq_2054

Query Match   22.0%;   Score 87.5;   DB 2;   Length 1116;
Best Local Similarity   56;   Conservative   35;   Mismatches   103;   Indels   61;   Gaps   9;

Qy   5   KNVIKEFMRFKVRMEGTVNGHEFEIEGEGE-----------------------GRPYEGHNTVKLKVTKGG 52
         |:|: | ::| | | : | :|||
Db   164   KDLLLDSEYFSAKTKGEILAEVEIKIRKNTGEILAEVEIKIRKENFTLSGTKINAKGTINLPVLDIN 223

RESULT 7
T06489
probable peptidylprolyl isomerase (EC 5.2.1.8) FKBP77 - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T06489
R;Brieman, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z15713
A;Accession: T06489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-568 <BRI>
A;Cross-references: UNIPROT:004843; UNIPARC:UPI000000AD075; EMBL:Y07636; PIDN:CAA68913.1
A;Experimental source: cv. ATIR, 2 day old plants, root tips
C;Genetics:
C;Gene: FKBP77
A;Superfamily: peptidylprolyl isomerase ROF1; BKBP-type peptidylprolyl isomerase homolog
C;Keywords: calmodulin binding; cis-trans-isomerase; cyclosporin A binding
F;54-101/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F;169-212/Domain: BKBP-type peptidylprolyl isomerase homology <PPI2>
F;286-334/Domain: BKBP-type peptidylprolyl isomerase homology <PPI1>
F;482-515/Domain: tetratricopeptide repeat homology <TTR>

Query Match   21.6%;   Score 87;   DB 2;   Length 568;
Best Local Similarity   61;   Conservative   32;   Mismatches   81;   Indels   108;   Gaps   15;

Qy   5   KNVIKE---------FMRFKVRMEGI-VNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPL 54
        :::| : || ||| | : |:| :| | ||| ::|
Db   39   KKLVKEGBGWPTAEETALKVEVHYTGTLLDGTKFD-SSRDRGTPF------KFKLBQGQVI 91

Qy   55   PFAWD-------------ILSPQFQYGSKVYVKHPADIPDYKKLSFP----------- 88
        | :| ||| :: | : | :||
Db   92   K-GWDQGIKTMKKGENASLTIPPDLAYGERA---PRTIPENATLRFDVELLSWASVKDI 146

Qy   89   ------------EGFKWERVNNFEDGGVVVTVTQDSSLQDGCFIYKVKFIGVNPFSPDGPVM 136
        |||| || | :|| | |:| ||
Db   147   CKDGGIFKKVLVEGGQKWB----NPKDLLDEVTVKYBARLEDGSVSKSESI---------- 192

Qy   137   QKKTMGWBASTERLYPRDGVLKGEIHKALK-LKDGGHYLVEFKSIY-------MAKKPVQLPG 191
       :| | :||| ||:| :| | : |:|| ||
Db   193   ------EFSV----KDGYFCPALSKAVKTMKKGEKVLLTVKPQYGFGEQGRAATEVEG 240

Qy   192   YYYVDSKLDI----------TSHNEDYTIV-------EQYER 216
       ::|| | | :| | :| ::||
Db   241   AVPPNSTLHIDLQLVSWKTLTLIGDDKRILKKVLKEGEGYER 282

RESULT 8
C97196
probable membrane protein [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: C97196

Qy   53   PLPFAWDILSPQFQYGS-----------------KVYVKHPADIPDYKKLSF------PEG---F 91
       ||| || |: |||
Db   224   AKAFVKDLIVRNKNYGSIEGIVKGNYELFPDLPLKGEAVNPEGTKIKFTYDVIPEGLLTF 283

Qy   92   KWER-VNNFEDGGVVVTVTQDSSLQDGCFIYKVKFIGVNPFSPDGPVMQKKTMGWEASTERL 150
       |:| | :|| : | || | :|
Db   284   SFENLVVDKNTLGINREIRGEFHGNGKVDFKKMFVKVN----------------AFTENL 327

Qy   151   YPRDGVLKGEIHKALKLKDGGHYLVIEFKSIYMAKKPVQLPGYYYVVD-SKLDITTSHNEDYT 209
       || | :||| :| : :||
Db   328   EVIDKKFKGDVLFSYNFSGNGSLNFEFKNSGYAK-----GNLIINKNKLEGBFSFNDPP 381

Db   210   IVEQ-----YERTEGR 220
Qy   382   VVFQDFNAYLSGBGK 396

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J;Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97196
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1484 <KUR>
A;Cross-references: UNIPROT:Q97GG4; UNIPARC:UPI00000CA4BD; GB:AB001437; PIDN:AAK80358.1;
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2403

Query Match   7.2%; Score 87; DB 2; Length 1484;
Best Local Similarity 20.7%; Pred. No. 33;
Matches 62; Conservative 47; Mismatches 76; Indels 114; Gaps 17;

```
Qy    9 KEFMRFKVRMEG----------TVNGHEFEIEG------------BGBGRPYEGHNTVKLKVTKGGP 53
Db  766 KNIVDFKTLLKKQNSFFVSENPREYKIEADIRDKDKNSNAPEVKGGIAEGDSANNWEVAKSGV 825

Qy   54 L-------PFAWDI-------LSPQFQ-------YGSKVYVKHPADIDYKKLSF-- 87
Db  826 LKARENNPYKFKINISGKNTNKNQHLKAKFYDANMHEIYSS--YIVKPSDSVDFESMDFTG 883

Qy   88 ---PEGFKWER--VNNFED--GGVVTVTQDSSLQDGGCFIYKVKFIGVNFPSDGPVMQKK 139
Db  884 ECVSPKTKYMRLEILSYKNPKGHKIYWKIHDVNIYD-LSEYKI-------SNSFVNNKK 934

Qy  140 TMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLV------------EFKSIYM 182
Db  935 VK--KAFTAKVIR-------ALVSKSGGKLKVTGKKSVYVDTFDESKSQFKNIYL 982

Qy  183 AKKPVQLPGYYYVDS-----KLD------ITSHNEDYTIVEQYERTEGRHHLF 224
Db  983 GK----FHFDSGKNSIKVENKQGFNAINLFTVIPENEYNTVTKIEKTADKCNIP 1033
```

RESULT 9
G69397
signal-transducing histidine kinase homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: G69397
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaec
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-456 <KLE>
A;Cross-references: UNIPROT:O29083; UNIPARC:UPI0000056B0A; GB:AE001022; GB:AE000782; NID:

Query Match   7.0%; Score 85.5; DB 2; Length 456;
Best Local Similarity 22.0%; Pred. No. 9.5;
Matches 46; Conservative 25; Mismatches 59; Indels 79; Gaps 10;

```
Qy   26 EFEIBGEBGEGRPYEGHHNTVKLKVTKGGPLPPAWDILSPQFQYGSKVYVKHPADIDYKKL 85
Db   67 EFEBLGE-EGRFYESR------YITKDGRVRHVWGPTAKTFLGKSYIIGNWIDVTKSKEL 120

Qy   86 SFPEGFKWERVMNFEDGGVVTVTQDS-----SLQDGCFIY------------KVKFIGVN 128
Db  121 -EQALRSSEEFYRTLVEDSLTPVYLLQDGIMVYVNKAFEEATGYKREEIVGRN 172

Qy  129 ----FPSDGPVMQKK 158
Db          -------------TMGWEASTERLYPRDGVLK------------------- 158
```

RESULT 10
F69442
hypothetical protein AF1543 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Accession: F69442
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-271 <KLE>
A;Cross-references: UNIPROT:O28729; UNIPARC:UPI0000056CB5; GB:AE000996; GB:AE000782; NID:
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1420

Query Match   7.0%; Score 85; DB 2; Length 271;
Best Local Similarity 25.6%; Pred. No. 5.2;
Matches 44; Conservative 18; Mismatches 64; Indels 46; Gaps 9;

```
Qy   64 QFQYGSKVYVKHP------ADIPDYKK--LSFPEGFKWERVMNFEDGGVV----TVTQDSSL 113
Db   65 QTEINSLFYTLHSLNLLGEDLPDYSSFLLKRLEGLKAERKYLLSDGGVTATYTFLQPNAL 124

Qy  114 QDGGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRD-----GVLKGEIHKALK 166
Db  125 RDA---YMISTLILHLYNRDVPEETKMLV---------------RRYRRETEFGVGYGVKKPNLEETFY 174

Qy  167 LKDGGHYLVEFKSIY-----------------MAKKPVQLPGY----YYVDSKLDITS 203
Db  175 ----ASYILRDKAVISFVKSFESNGGFAKQPGGYYPPYLEDTYYATSTLSLLS 222
```

RESULT 11
A40351
adhesion-type protein ADMLX - human
C;Species: Homo sapiens (man)
C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 15-Mar-2004
C;Accession: A40351
R;Legouis, R.; Hardelin, J.P.; Levilliers, J.; Claverie, J.M.; Compain, S.; Wunderle, V.;
, G.; Weissenbach, J.; Petit, C.
Cell 67, 423-435, 1991
A;Title: The candidate gene for the X-linked Kallmann syndrome encodes a protein related
A;Reference number: A40351; MUID:92005720; PMID:1913827
A;Accession: A40351
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-679 <LBG>
A;Cross-references: UNIPARC:UPI000016B3D2; GB:S60085; NID:g2375595; PIDN:AAB20108.1; PID:S
A;Gene: GDB:KAL1; KAL
C;Genetics:
A;Gene: GDB:120116; OMIM:308700
A;Map position: Xp22.32-Xp22.32
F;129-175/Domain: antileukoproteinase repeat homology <ALP>

Query Match   7.0%; Score 84.5; DB 2; Length 679;
Best Local Similarity 27.0%; Pred. No. 20;
Matches 48; Conservative 20; Mismatches 75; Indels 35; Gaps 10;

```
Qy   46 LKVTKGG---PLPF-------AWDILSPQFQYGS----KVYVKHPAD--IPDYKKLSFPEG 90
Db  173 PFFLIHPEDRGLVYKRYIEREKGLRDIMETYSW-----RIIRKDGEVRWVTARPGRVTYR 227

Qy  159 -------GEIHKA-LKLKDGGHYL 174
Db  228 GRPAVAATVVDTEIHKLNIELKKGEYL 256
```

```
Db   406 VKTRKGGIQTQLPFQRRRPTRPLEVGAPFYQDGQLQVKRVWKKTEDPTVNRYHVRWFPEA 465
              |  |||     |  |  :  |:                            :||| |::|
Qy    91 FKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERL 150

Db   466 CAHNRTTGSEASSGWTHENYIILQDLSFSCKYKV-------TVQPIRPKS---HSKAEAV 515
              :|                |:                       ||| ::      ||
Qy   151 Y-----PRDGVLKGEIHKALK-LKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSH 204

Db   516 FFTTPPCSALKGKSHKPIGCLGEAGHVLSK-----VLAKPENLSASFIVQD-VNITGH 567
```

RESULT 12

IJMSNG
neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse
N;Alternate names: NCAM-120
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C;Accession: A29673; S00382; A44290
R;Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec
EMBO J. 6, 907-914, 1987
A;Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000
A;Reference number: A29673; MUID:87246524; PMID:3595563
A;Molecule type: mRNA
A;Accession: A29673
A;Residues: 1-725 <BAR>
A;Cross-references: UNIPROT:P13594; UNIPARC:UPI000002BAC3; EMBL:Y00051; NID:g53342; PIDN
R;Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A;Title: Differential splicing and alternative polyadenylation generates distinct NCAM t
A;Reference number: S00382; MUID:88283628; PMID:3396534
A;Accession: S00382
A;Molecule type: DNA
A;Residues: 642-656,'D',658-725 <BA2>
A;Cross-references: UNIPARC:UPI000174377; EMBL:X07195
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A;Title: Structural and immunological characterization of the amino-terminal domain of m
A;Reference number: A44290; MUID:86140120; PMID:3512556
A;Accession: A44290
A;Residues: 20-36 <ROU>
A;Cross-references: UNIPARC:UPI000174376
A;Comment: NCAM mediates cell-cell adhesion via a homophilic binding with another NCAM mol
C;Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IJMS
C;Genetics:
A;Gene: NCAM
A;Map position: 9
A;Introns: 701/1
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane
F;1-19/Domain: signal sequence #status predicted <SIG>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN3A>
F;625-685/Domain: fibronectin type III repeat homology <FN3B>
F;41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted
F;222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   7.0%; Score 84.5; DB 1; Length 725;
Best Local Similarity   24.2%; Pred. No. 21;
Matches 37; Conservative 30; Mismatches 57; Indels 29; Gaps 9;

```
Qy    91 FKWERVMNFEDGGVVTV---TQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEAST 147
Db   548 FTWYDAKEANMEGIVTIMGLKPETTYSD-------RLAALNGKGLGBIMQPSESKTQPVP 600
Qy   148 ERLYPRDGVLKGEIHK--------ALKLKDGG---HYLVEFKSIYMAKKPVQLPGYY 193
```

```
Db   601 ELSAPK---LEGQMGEDGNSIKVNLIKQDDGSPIRIYLVKYRALASEWKPEIRLPSGSH 657
Qy   194 YVDSK-LDITSHNEDYTIVE-QYERTEGRHHLP 224
Db   658 HVMLKSIDWNABYEVVVAENOQGKSKAAHFVP 690
```

RESULT 13

T17884
S-layer protein - Bacillus circulans
C;Species: Bacillus circulans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T17884
R;Aubert-Pivert, E.; Davies, J.
Gene 147, 1-11, 1994
A;Title: Biosynthesis of butirosin in Bacillus circulans NRRL B3312: identification by se
A;Reference number: Z18808; MUID:94374689; PMID:7522196
A;Residues: 1-1616 <AUB>
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Cross-references: UNIPROT:P35824; UNIPARC:UPI0000135A00; EMBL:L20421; NID:g304142; PID:
C;Genetics:
A;Gene: butB
C;Function:
A;Pathway: butirosin biosynthesis

Query Match   6.9%; Score 84; DB 2; Length 1616;
Best Local Similarity   20.6%; Pred. No. 69;
Matches 59; Conservative 48; Mismatches 85; Indels 94; Gaps 17;

```
Qy     6 NVIKEFMR----------------FKVRMEGTVNGHEFEI---EGEGEGRPYEGHNTVKLKVTK 50
Db   590 DILQEFIRYSRELGLIDIHVSFNIFAEGSIASNEFALLDSHLDWEERVVTNAADNGQIK--- 646
Qy    51 GGPLPFAWDILSPQFQYGSKVYVKHPAD-IPDYKKLSFPEGFKWERVMNFEDGGVV--TV 107
Db   647 -----RLRESAKQGAVAFVNPSNDEVRDFQLKTIB-----EVLQNYDVDGVVLDRA 692
Qy   108 TQDSSLQDGCFIYKVK----FIGV------NFP--------SDGPVMQ-------KKT 140
Db   693 RYDNESADFSDLTKAKPESFLGARGKQLQNWFDDVFTYAGNVRKDGPLIRDWFPRSKTI 752
Qy   141 MGWEASTERLYPRDGVLKGEIHKALKLK-----------DGGHY-LVEFK---SIYMAK 184
Db   753 KSFTSEVRQLTDR---VKAEKGKKIEVSAYVGSWFESYYLNGVHWGSTEFRYDERLRMKD 809
Qy   185 KPVQLPGYY---YVDSKLDITSHNEDYTIVEQYERT--EGRHHLFL 225
Db   810 KSVYTPGYYESGYV---------KNLDFIMIGAYQTTAPEIEHYITL 847
```

RESULT 14

E90276
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: E90276
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-W
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: E90276
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-268 <KUR>
A;Cross-references: UNIPROT:Q97YT6; UNIPARC:UPI000006438A; GB:AE006641; NID:g13814420; P)
C;Genetics:
A;Gene: SSO1221

```
RESULT 15

E69544
hypothetical protein AF2357 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: E69544
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso
R;Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
. Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: E69544
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-340 <KLE>
A;Cross-references: UNIPROT:O30313; UNIPARC:UPI00000572DC; GB:AE001113; GB:AE000782; NID

Query Match      6.9%;  Score 83.5;  DB 2;  Length 340;
Best Local Similarity   23.9%;  Pred. No. 9.6;
Matches   38;  Conservative   29;  Mismatches   63;  Indels   29;  Gaps    9;

Qy    41 HNTVKLKVTKGGPLPFAWDILSPQFQYGSKVYKHPADIPDYKKLSPPEGFKWERVMNFE 100
            ||::|         ||||||||||||||   : |::|||  ||||||||||||||||
Db   144 HAWEVKINN------SWVVADPTVYW---LYVNYPEKYPNWNKLWP-NNESWANLIDF- 192

Qy   101 DGGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQ--KKTMGWEASTERLYPRDGVLK 158
            ::||| ||||||||||||||||||||||||||||   ||||||||||||||||||||
Db   193 -SRVTVLPNGSVLDLTSNY-TKTYNVTITIDQNVEKGILKVTTWKGSVERTVYSKAVNK 250

Qy   159 GE-IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVD 196
            || :|| || :||||||||||: ||||||:| | :|||
Db   251 SDTVNLALATR-----IYKFELI-------VPTWYFLE 276


Query Match      6.9%;  Score 83.5;  DB 2;  Length 268;
Best Local Similarity   25.8%;  Pred. No. 7;
Matches   49;  Conservative   21;  Mismatches   59;  Indels   61;  Gaps   13;

Qy    34 EGRPYEGHNTVKLKVTKGG-------PLPFAWDILSPQFQYGSKVY---------VKHPAD 78
            |||| |    |    ||                    |     |:           |:
Db     3 EIRPY--YNTFSLPITKSGKSQIVPPPWIVAI-----EMIGVKAYFDPVKVLDLVPPPLE 56

Qy    79 IPD----YKKLSPPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF---IGVNF 129
            |      ||:|          :|:   ::|||: |||||||||||||||     :||:
Db    57 IVDGEGFVYIAKIFTVSGNRWE--MLYED------PEETKYMEAAIALKVKVNDNIFTYF 108

Qy   130 P-----SDGPVMQKKTMGWEASTERLYPRD--GVLKGEIHKALK--------LKDGGHYL 174
            |      |||| :|  :| |||||||||||    :||:||:|            ||||||
Db   109 PFMWVDKDLPLLR----GWLLG---YPKKLAYISISBFHKLLDGYSGPSSGVRMGGYAL 160

Qy   175 VEFKSIYMAK 184
            :  :|::|| |
Db   161 RNGKEIIRVK 170
```

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

Copyright (c) 1993 - 2006 Compugen Ltd.
GenCore version 5.1.6

Run on: January 11, 2006, 01:54:16 ; Search time 73 Seconds
(without alignments)
2174.574 Million cell updates/sec

Title: US-10-006-922A-12
Perfect Score: 1214
Sequence: 1 MRSSKNVIKEPMRPKVRMEG.........EDYIIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   UniProt_05.80:*
              1: uniprot_sprot:*
              2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	*Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	225	1	RPP_DISSP	Q9u6y8 discosoma s
2	1196	98.5	236	2	Q66ND7_9CNID	Q66nd7 discosoma s
3	1193	98.3	225	2	Q6KF85_9CNID	Q6kf85 discosoma s
4	1191	98.1	236	2	Q66ND8_9CNID	Q66nd8 discosoma s
5	1085.5	89.4	230	2	Q9GTJ7_9CNID	Q9gtj7 discosoma s
6	985	81.1	226	2	Q53G08_9CNID	Q53g08 discosoma s
7	972	80.1	226	2	Q53G07_9CNID	Q53g07 discosoma s
8	804	66.2	221	2	Q66PV1_ACRTE	Q66pv1 acropora te
9	798	65.7	221	2	Q66PU8_9CNID	Q66pu8 acropora ac
10	797	65.7	221	2	Q66PV0_ACRMI	Q66pv0 acropora mi
11	793	65.3	221	1	NFCP_GONTE	Q95p04 goniopora t
12	793	65.3	221	1	NFCP_MONEF	P83690 montipora e
13	782	64.4	221	1	Q66ND6_DISST	Q66nd6 discosoma s
14	745	61.4	227	1	GFPL_DISST	GPPL_DISST discosoma s
15	729.5	60.1	232	1	Q9U6Y7_9CNID	Q9u6y7 discosoma s
16	690	56.8	225	2	Q963F5_9CNID	Q963f5 discosoma s
17	684	56.3	225	2	Q66ND3_9CNID	Q66nd3 montastraea
18	681	56.1	225	2	Q7Z0W4_9CNID	Q7z0w4 montastraea
19	681	56.1	225	2	Q95UA7_9CNID	Q95ua7 montastraea
20	681	56.0	227	2	Q72OW6_9CNID	Q72ow6 montastraea
21	679.5	56.0	227	2	Q72OW8_9CNID	Q72ow8 montastraea
22	678.5	55.9	227	2	Q962P9_9CNID	Q962p9 montastraea
23	678	55.8	225	2	Q72OW5_9CNID	Q72ow5 montastraea
24	678	55.8	224	2	Q8TLG6_9CNID	Q8tlg6 echinophyll
25	677	55.8	227	2	Q66ND5_9CNID	Q66nd5 montastraea
26	672.5	55.4	227	2	Q66ND2_9CNID	Q66nd2 montastraea
27	670.5	55.2	227	2	Q66ND5_9CNID	Q66nd5 montastraea
28	670.5	55.1	234	2	Q7Z0W7_9CNID	Q7z0w7 montastraea
29	668.5	55.1	234	2	Q8T5F2_9CNID	Q8t5f2 montastraea
30	668.5	55.0	225	2	Q6OI25_9CNID	Q6oi25 montastraea
31	667	54.9	226	2	Q6OI24_9CNID	Q6oi24 montastraea
32	667	54.9	266	1	GPPL_CLASP	Q9u6y3 clavularia
33	663	54.6	226	2	Q8T6U0_9CNID	Q8t6u0 dendronepht
32	660.5	54.4	226	2	Q66ND4_9CNID	Q66nd4 montastraea
33	655.5	54.0	226	2	Q66PU5_9CNID	Q66pu5 agaricia fr
34	655.5	54.0	227	2	Q95VT0_9CNID	Q95vt0 montastraea
35	654.5	53.9	234	2	Q8MU47_9CNID	Q8mu47 montastraea
36	652.5	53.7	225	2	Q7Z0W9_9CNID	Q7z0w9 montastraea
37	649.5	53.5	227	2	Q53UG7_9CNID	Q53ug7 favia favus
38	649	53.5	225	2	Q60I24_9CNID	Q60i24 galaxea fas
39	642.5	53.0	225	2	Q6USK3_9CNID	Q6usk3 montastraea
40	639.5	52.9	227	2	Q53UG8_9CNID	Q53ug8 favia favus
41	639	52.7	224	2	Q8MU48_9CNID	Q8mu48 montastraea
42	631.5	52.1	225	2	Q8T5F1_9CNID	Q8t5f1 montastraea
43	620.5	51.1	230	2	Q66PW1_9CNID	Q66pw1 scolymia cu
44	610	50.2	223	2	Q68BF5_9CNID	Q68bf5 astrangia 1
45	608.5	50.1	231	1	RPP_PARAC	Q81sf8 parasicyoni

ALIGNMENTS

RESULT 1
RPP_DISSP
ID   RPP_DISSP   STANDARD;   PRT;   225 AA.
AC   Q9U6Y8;
DT   01-FEB-2005 (Rel. 46, Created)
DT   01-FEB-2005 (Rel. 46, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Red fluorescent protein drFP583 (DsRed).
OS   Discosoma sp. (Sea anemone).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=86600;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=99436614; PubMed=10504696; DOI=10.1038/13657;
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RT   "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL   Nat. Biotechnol. 17:969-973(1999).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   PubMed=11101896; DOI=10.1038/81992;
RA   Wall M.A., Socolich M., Ranganathan R.;
RT   "The structural basis for red fluorescence in the tetrameric GFP
RT   homolog DsRed.";
RL   Nat. Struct. Biol. 7:1133-1138(2000).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND SUBUNIT.
RX   PubMed=11209050; DOI=10.1073/pnas.98.2.462;
RA   Yarbrough D., Wachter R.M., Kallio K., Matz M.V., Remington S.J.;
RT   "Refined crystal structure of DsRed, a red fluorescent protein from
RT   coral, at 2.0-A resolution.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:462-467(2001).
-!- FUNCTION: Thought to play a role in photoprotection of the coral's
    resident symbiont microalgae's photosystems from photoinhibition
    caused by high light levels found near the surface of coral reefs.
    In deeper water, the fluorescence may be to convert blue light
    into longer wavelengths more suitable for use in photosynthesis by
    the microalgal symbionts.
-!- BIOPHYSICOCHEMICAL PROPERTIES:
    Absorption:
      Abs(max)=558 nm;
      Note=Exhibits a smaller absorbance peak at 494 nm. The broad
      fluorescence emission spectrum peaks at 583 nm;
-!- SUBUNIT: Homotetramer.
-!- PTM: Contains a chromophore consisting of modified amino acid
    residues. The chromophore is formed by autocatalytic backbone
    condensation between Xaa-N and Gly-(N+2), oxidation of Tyr-(N+1)
    to didehydrotyrosine, and formation of a double bond to the alpha-
    amino nitrogen of residue Xaa-N. Maturation of the chromophore
    requires nothing other than molecular oxygen.
-!- BIOTECHNOLOGY: Fluorescent proteins have become a useful and
    ubiquitous tool for making chimeric proteins, where they function
    as a fluorescent protein tag. Typically they tolerate N- and C-

```
CC   terminal fusion to a broad variety of proteins. They have been
CC   expressed in most known cell types and are used as a noninvasive
CC   fluorescent marker in living cells and organisms. They enable a
CC   wide range of applications where they have functioned as a cell
CC   lineage tracer, reporter of gene expression, or as a measure of
CC   protein-protein interactions.
CC   -!- SIMILARITY: Belongs to the GFP family.
CC   -----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   -----------------------------------------------------------------------
DR   EMBL; AF168419; AAF03369.1; -; mRNA.
DR   PDB; 1G7K; X-ray; A/B/C/D=2-225.
DR   PDB; 1GGX; X-ray; A/B/D=1-225.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFPLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
KW   3D-structure; Chromophore; Luminescence; Photoprotein.
FT   MOD_RES       67     67      (Z)-2,3-didehydrotyrosine.
FT   CROSSLNK      66     68      2-iminomethyl-5-imidazolinone (Gln-Gly).
SQ   SEQUENCE   225 AA;  25931 MW;  FBFPA569778F689 CRC64;

Query Match             100.0%;  Score 1214;  DB 1;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-96;
Matches  225;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 2
Q66ND7_9CNID
ID   Q66ND7_9CNID PRELIMINARY;      PRT;   236 AA.
AC   Q66ND7;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Enhanced red fluorescent protein R+.
OS   Discosoma sp. RC-2004.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=289055;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT   "Cloning of Anthozoan Fluorescent Protein Genes.";
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY679107; AAU04444.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFPLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   236 AA;  27032 MW;  BB3B4BE06B29EF0 CRC64;

Query Match             98.3%;  Score 1193;  DB 2;  Length 225;
Best Local Similarity   98.2%;  Pred. No. 6.7e-95;
Matches  221;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 3
Q6KF85_9CNID
ID   Q6KF85_9CNID PRELIMINARY;      PRT;   225 AA.
AC   Q6KF85;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Orange fluorescent protein FP586.
OS   Discosoma sp. JW-2002.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Wiedenmann J., Girod A.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF545828; AAQ11987.1; -; mRNA.
DR   HSSP; P42212; 1B9C.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFPLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   225 AA;  25791 MW;  E151D0E497AA23FA CRC64;

Query Match             98.2%;  Score 1196;  DB 2;  Length 225;
Best Local Similarity   98.2%;  Pred. No. 3.9e-95;
Matches  221;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|:
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|:
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 4
Q66ND8_9CNID
ID   Q66ND8_9CNID PRELIMINARY;      PRT;   236 AA.
AC   Q66ND8;
```

```
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Red fluorescent protein R1.
OS   Discosoma sp. RC-2004.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=289055;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT   "Cloning of Anthozoan Fluorescent Protein Genes.";
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY679106; AAU04443.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE  236 AA;  27042 MW;  BB3844BCD6834EF3 CRC64;

Query Match             98.1%;  Score 1191;  DB 2;  Length 236;
Best Local Similarity   97.8%;  Pred. No. 1.1e-94;
Matches 220;  Conservative   2;  Mismatches   3;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db          1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDPSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||:|
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 5
ID   Q9GTJ7_9CNID  PRELIMINARY;     PRT;   230 AA.
AC   Q9GTJ7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Red fluorescent protein.
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA   Lukyanov S.A.;
RT   "Novel fluorescent protein from Discosoma coral and its mutants
RT   possesses a unique far-red fluorescence.";
RL   FEBS Lett. 479:127-130(2000).
EMBL; AF272711; AAG16224.1; -; mRNA.
HSSP; Q9U6Y8; 1GGX.
OS   Discosoma sp. SSAL-2000.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
GN   Name=FP593;
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein.
SQ   SEQUENCE  230 AA;  26370 MW;  5215B1B436D67B51 CRC64;

Query Match             89.4%;  Score 1085.5;  DB 2;  Length 230;
Best Local Similarity   88.1%;  Pred. No. 1.3e-85;
Matches 199;  Conservative  17;  Mismatches   9;  Indels   1;  Gaps   1;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||:||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db          1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||:|:||||||||||:||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        121 EVKFIGVNFPSDGPVMQRRTRGWEASSRLYPRDGVLKGDIHWALRLEGGGHYIVEFKSI 180

Qy        181 YMAKKP-VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||   :|||||||||||||:||||:|||||||||||||| |:
Db        181 YMVKKPSVQLPGYYYVDSKLDMTSHNEDYTIVEQYEKTQGRHHPPI 226
```

```
RESULT 6
ID   Q5S3G8_9CNID  PRELIMINARY;     PRT;   226 AA.
AC   Q5S3G8;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Fluorescent protein raspberry.
OS   Discosoma sp. LW-2004.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=301246;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Wang L., Jackson W.C., Steinbach P.A., Tsien R.Y.;
RX   PubMed=15556995; DOI=10.1073/pnas.0407752101;
RT   "Evolution of new nonantibody proteins via iterative somatic
RT   hypermutation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:16745-16749(2004).
DR   EMBL; AY786536; AAV65486.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE  226 AA;  25513 MW;  159A8FB82D453A5B CRC64;

Query Match             81.1%;  Score 985;  DB 2;  Length 226;
Best Local Similarity   85.3%;  Pred. No. 6.3e-77;
Matches 185;  Conservative   6;  Mismatches  26;  Indels   0;  Gaps   0;

Qy          5 KNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQ 64
              :||||||||||||||||||||||||||||||||:|||||   ||||||||||||:||||
Db          6 EEVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ 65

Qy         65 FQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF 124
              :||||||||||||||||||||||||||||||:|||| |||||||||||||||||:|||||
Db         66 CMYGSKGYVKHPADIPDYKKLSFPEGFKWERVMNFEDGALKGEMKVRLKLKDGGHYDAEVKF 125

Qy        125 IGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 184
              |||||||||||||||:|||||:||||||||||||||||||||||||||||||:||||||
Db        126 RGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEMKVRLKLKDGGHYDAEVKTTYMAK 185

Qy        185 KPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
              ||||||||||||||||||||||||||||||||||||
Db        186 KPVQLPGAYKYDIKLDITSHNEDYTIVEQYERABGRH 222
```

```
RESULT 7
ID   Q5S3G7_9CNID  PRELIMINARY;     PRT;   226 AA.
```

AC   Q5S3G7;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Fluorescent protein plum.
OS   Discosoma sp. LW-2004.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=301246;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15556995; DOI=10.1073/pnas.0407752101;
RA   Wang L., Jackson W.C., Steinbach P.A., Tsien R.Y.;
RT   "Evolution of new nonantibody proteins via iterative somatic
RT   hypermutation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:16745-16749(2004).
DR   EMBL; AY786537; AAV65487.1; -; mRNA.
SMR; Q5S3G7; 8-222.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR000786; GFP_related.
DR   InterPro; IPR011584; GFP_related.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   226 AA; 25590 MW; 6DF38CAD2AB2BBED CRC64;

Query Match   80.1%; Score 972; DB 2; Length 226;
Best Local Similarity   83.4%; Pred. No. 8.4e-76;
Matches 181; Conservative   8; Mismatches 28; Indels   0; Gaps   0;

Qy   5 KNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQ 64
          ||||||||||||||||||||||||||||||||||||| |||||||| |||||||||||||
Db   6 EEVIKEFMRFKFKEHMEGSVNGHEFEIEGEGEGRPYEGTQTOTAARLKVTKGGPLPFAWDILSPQ 65

Qy   65 FQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF 124
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 IMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKV 125

Qy   125 IGVNFPSDGPVWMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 184
          ||||||||||||||||||||||||||||||||||||||||| |||| | | | ||||
Db   126 RGTNFPSDGPVWMQKKTMGWEASSERMYPEDGALKGEMKMRLRLKDGGHYDAEVKTTYMAK 185

Qy   185 KPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
          |||||||||||||||||||||||||||||| ||||
Db   186 KPVQLPGAYKDIKLDITSHNEDYTIVEQYERAEGGRH 222

RESULT 8
ID   Q66PV1_ACRTE
AC   Q66PV1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Alieva N.O., Meleshkevitch E.A., Field S.F., Matz M.V.;
RT   "Survey of coral GRP-like proteins.";
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
EMBL; AY646074; AAU06853.1; -; mRNA.
GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
InterPro; IPR011584; GFP_related.
Pfam; PF01353; GFP; 1.
PRINTS; PR01229; GFLUORESCENT.
ProDom; PD013756; Green_fl_protein; 1.
NCBI_TaxID=70783;
OS   Acropora tenuis (Purple tipped acroporal).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Astrocoeniina; Acroporidae; Acropora.
Chromoprotein

Query Match   66.2%; Score 804; DB 2; Length 221;
Best Local Similarity   67.9%; Pred. No. 2.7e-61;
Matches 148; Conservative   25; Mismatches 45; Indels   0; Gaps   0;

Qy   6 NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF 65
          |||| ||||||||||||||||||||||||||||||||::||:|:||||||||||||||||
Db   2 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQF 61

Qy   66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125
          |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db   62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDDAVCTVSNDSSIQGNCFIYHVKFS 121

Qy   126 GVNFPSDGPVWMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 185
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   122 GVNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy   186 PVRMPGYHYVDYKRLDVTNHHRDYTSVSEQCRISIARKPL 219
          |||||||:|||:|| :|| ||:|| :|| |
Db   182 PVRMPGYHIYVDRKLDVTNHNDYTSVEQCRISIARKPL 219

RESULT 9
ID   Q66PU8_9CNID
AC   Q66PU8;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Acropora aculeus.
OS   Acropora aculeus.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Astrocoeniina; Acroporidae; Acropora.
NCBI_TaxID=287157;
RN   [1]
NUCLEOTIDE SEQUENCE.
RA   Alieva N.O., Meleshkevitch E.A., Field S.F., Matz M.V.;
RT   "Survey of coral GRP-like proteins.";
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
EMBL; AY646077; AAU06856.1; -; mRNA.
GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
InterPro; IPR011584; GFP_related.
Pfam; PF01353; GFP; 1.
PRINTS; PR01229; GFLUORESCENT.
ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   221 AA; 24976 MW; 307DE4FC4B01B843 CRC64;

Qy   6 NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF 65
          |||| ||||||||||||||||||||||||||||||::||:|:||||||||||||||||||
Db   2 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQS 61

Qy   66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125
          |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db   62 QYGSIPFTKYPDDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYNVKFS 121

Qy   126 GVNFPSDGPVWMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 185
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy   186 PVQLPGYYYVDSKLDITSHNEDYTIVEQYE 215
          ||:||||||||:|||:|||||||||||||
Db   182 PVRMPGHIYVDRKLDVTNHNRDYTSVEQCE 211

RESULT 10
ID   Q66PV0_ACRMI
AC   Q66PV0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)

```
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Chromoprotein.
OS   Acropora millepora (Coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Astrocoeniina; Acroporidae; Acropora.
OX   NCBI_TaxID=45264;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Alieva N.O., Meleshkevitch E.A., Field S.P., Matz M.V.;
RT   "Survey of coral GFP-like proteins.";
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY646075; AAU06854.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFPLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   221 AA;  24987 MW;  546D2CFPAF25522F2 CRC64;

Query Match             65.7%;  Score 797;  DB 2;  Length 221;
Best Local Similarity   68.6%;  Pred. No. 1.1e-60;
Matches  144;  Conservative  26;  Mismatches  40;  Indels  0;  Gaps  0;

Qy      6 NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF 65
          |||||||||: :|||||:||||||||||||||||| |||||||||:||||||||||||||
Db      2 SVIAKQMTYKVYVMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQC 61

Qy     66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125
          ||||| ||||| ||||||| ||||| ||||| |||||||||||||||||||||||||: |
Db     62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy    126 GVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKK 185
          ||||| |||||||||| ||||||||||||| ||||:|||||:|||||| |||||||||||
Db    122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLLGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy    186 PVQLPGYYYVDSKLDITSHNEDYTIVEQYE 215
          |||||||||||| |||:|||||||:||||||
Db    182 PVKMPGYHYVDRKLDVTNNKDYTSVEQCE 211
```

```
RESULT 11
NFCP_GONTE
ID   NFCP_GONTE    STANDARD;    PRT;    221 AA.
AC   Q95P04;
DT   01-FEB-2005 (Rel. 46, Created)
DT   01-FEB-2005 (Rel. 46, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   GFP-like non-fluorescent chromoprotein (gtCP).
OS   Goniopora tenuidens (Coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Fungiina; Poritidae; Goniopora.
OX   NCBI_TaxID=75301;
RN   [1]
RP   NUCLEOTIDE SEQUENCE, SUBUNIT, AND MUTAGENESIS OF ASN-142.
RX   MEDLINE=21538626; PubMed=11682051; DOI=10.1016/S0014-5793(01)02930-1;
RA   Gurskaya N.G., Fradkov A.F., Terskikh A., Matz M.V., Labas Y.A.,
RA   Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT   "GFP-like chromoproteins as a source of far-red fluorescent
RT   proteins.";
RL   FEBS Lett. 507:16-20(2001).
-!- FUNCTION: Non-fluorescent pigment protein that is lilac in color.
-!- BIOPHYSICOCHEMICAL PROPERTIES:
      Absorption:
        Abs(max)=580 nm;
-!- SUBUNIT: Homotetramer.
-!- PTM: Contains a chromophore consisting of modified amino acid
    residues. The chromophore is formed by autocatalytic backbone
    condensation between Xaa-N and Gly-(N+2), oxidation of Tyr-(N+1)
    to didehydrotyrosine, and formation of a double bond to the alpha-
    amino nitrogen of residue Xaa-N. Maturation of the chromophore
    requires nothing other than molecular oxygen. The precise
```

```
-!- MISCELLANEOUS: Fluorescence excitation of the Ser-142 mutant is at
    589 nm and emission at 615 nm.
-!- SIMILARITY: Belongs to the GFP family.

DR   EMBL; AF383156; AAL27542.1; -; mRNA.
DR   HSSP; Q9U6Y8; 1GGX.
DR   GO; GO:0018995; C:host; ISS.
DR   GO; GO:0010117; P:photoprotection; ISS.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFPLUORESCENT.
KW   Chromophore; Luminescence; Photoprotein.
FT   MOD_RES      63     63       2,3-didehydrotyrosine (By similarity).
FT   CROSSLNK     62     64       2-iminomethyl-5-imidazolinone (Gln-Gly)
FT                               (By similarity)
FT                               N->S: Produces a fluorescent form.
FT   MUTAGEN     142    142
SQ   SEQUENCE   221 AA;  24918 MW;  93F9F4B5C2003CB4 CRC64;

Query Match             65.3%;  Score 793;  DB 1;  Length 221;
Best Local Similarity   68.6%;  Pred. No. 2.4e-60;
Matches  144;  Conservative  25;  Mismatches  41;  Indels  0;  Gaps  0;

Qy      6 NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF 65
          |||||||||: :|||||:||||||||||||||||| |||||||||:||||||||||||||
Db      2 SVIAKQMTYKVYVMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQS 61

Qy     66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125
          ||||| ||||| ||||||| ||||| ||||| |||||||||||||||||||||||||: |
Db     62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy    126 GVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKK 185
          ||||| |||||||||| ||||||||||||| ||||:|||||:|||||| |||||||||||
Db    122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy    186 PVQLPGYYYVDSKLDITSHNEDYTIVEQYE 215
          |||||||||||| |||:|||||||:||||||
Db    182 PVKMPGYHYVDRKLDVTNNIDYTSVEQCE 211
```

```
RESULT 12
Q66PU9_9CNID
ID   Q66PU9_9CNID PRELIMINARY;    PRT;    221 AA.
AC   Q66PU9;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Chromoprotein.
OS   Acropora hyacinthus.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Astrocoeniina; Acroporidae; Acropora.
OX   NCBI_TaxID=55974;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Alieva N.O., Meleshkevitch E.A., Field S.P., Matz M.V.;
RT   "Survey of coral GFP-like proteins.";
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
```

Left column

DR   EMBL; AY646076; AAU06855.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   221 AA;   24963 MW;   FC9D9749069D079 CRC64;

RESULT 13
ID   NFCP_MONEF   STANDARD;   PRT;   221 AA.
AC   P83690;
DT   01-FEB-2005 (Rel. 46, Created)
DT   01-FEB-2005 (Rel. 46, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   GFP-like non-fluorescent chromoprotein (Rtms 5) (Non-fluorescent
DE   pocilloporin).
OS   Montipora efflorescens (coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Astrocoeniina; Acroporidae; Montipora.
OX   NCBI_TaxID=105610;
RN   [2]
RP   CRYSTALLIZATION.
RX   PubMed=12595737; DOI=10.1107/S0907444902023466;
RA   Beddoe T., Ling M., Dove S., Hoegh-Guldberg O., Devenish R.J.,
RA   Prescott M., Rossjohn J.;
RT   "The production, purification and crystallization of a pocilloporin
RT   pigment from a reef-forming coral.";
RL   Acta Crystallogr. D 59:597-599(2003).

Query Match   68.4%; Score 793; DB 2; Length 221;
Best Local Similarity   65.3%;
Matches 144; Conservative 25; Mismatches 41; Indels   0; Gaps   0;

Qy   6 NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF 65
              |::|||:|||||:|:|:||:|:|||||||||||:||:|||:||||||||||||||||||
Db   2 SVIATQMTYKVYMSGTVNGHYFEVEGDGKGRPYEGEQTVRLTVTKGGPLPFAWDILSPQF 61

Qy   66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125
              ::|:||:|||||||||:| ||||||||:|||:|||||||:|||:||||||:|||| ||
Db   62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy   126 GVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKK 185
              ||||:||||:||||:|||||||||||||||:|||:|||:|||||||||:|||:|||||
Db   122 GLNFPPNGPVMQKKTQGWEPHSERLFARDGVLIGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy   186 PVQLPGYYYVDSKLDITSHNEDYTIVEQYE 215
              ||:||||||||:|||||||:||||||||||
Db   182 PVKMPGYHFVDRKLDVTNHNKDYTSVEQRE 211

Center column

CC   expressed in most known cell types and are used as a noninvasive
CC   fluorescent marker in living cells and organisms. They enable a
CC   wide range of applications where they have functioned as a cell
CC   lineage tracer, reporter of gene expression, or as a measure of
CC   protein-protein interactions.
CC   -!- MISCELLANEOUS: The wild-type form is non-fluorescent with the
CC   color being pH dependent, ranging from yellow at low pH, through
CC   red to blue at high pH. The His-142 mutation produces a
CC   fluorescent form.
CC   -!- SIMILARITY: Belongs to the GFP family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --

RESULT 14
ID   Q66ND6_DIST   Q66ND6_DIST PRELIMINARY;   PRT;   227 AA.
AC   Q66ND6;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Cyan fluorescent protein C1.
OS   Discosoma striata (Striped mushroom).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=105400;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT   "Cloning of Cnidarian Fluorescent Protein Genes.";
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY679108; AAU04445.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.

Query Match   68.1%; Score 782; DB 1; Length 221;
Best Local Similarity   64.4%;
Matches 143; Conservative 23; Mismatches 44; Indels   0; Gaps   0;

Qy   6 NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQC 65
              |::|||:|||||:|:|:||:|:|||||||||||:||:|||:||||||||||||||||||:
Db   2 SVIATQMTYKVYMSGTVNGHYFEVEGDGKGRPYEGEQTVKLTVTKGGPLPFAWDILSPQC 61

Qy   66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125
              ::|:||:|||||||||:| ||||||||:|||:||||||||:|||:||||||:|||| ||
Db   62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy   126 GVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKK 185
              ||||:||||:||||:|||||||||||||||:|||:|||:|||||||||:|||:|||||
Db   122 GLNFPPNGPVMQKKTQGWEPHSERLFARDGVLIGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy   186 PVQLPGYYYVDSKLDITSHNEDYTIVEQYE 215
              ||:||||:|||:|||||||:||||||||||:
Db   182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 211

Right column

-!- FUNCTION: Thought to play a role in photoprotection of the coral's
    resident symbiont microalgae's photosystems from photoinhibition
    caused by high light levels found near the surface of coral reefs.
-!- SUBUNIT: Homotetramer.
-!- PTM: Contains a chromophore consisting of modified amino acid
    residues. The chromophore is formed by autocatalytic backbone
    condensation between Xaa-N and Gly-(N+2), oxidation of Tyr-(N+1)
    to didehydrotyrosine, and formation of a double bond to the alpha-
    amino nitrogen of residue Xaa-N. Maturation of the chromophore
    requires nothing other than molecular oxygen.
-!- BIOTECHNOLOGY: Fluorescent proteins have become a useful and
    ubiquitous tool for making chimeric proteins, where they function
    as a fluorescent protein fusion to a broad variety of proteins.
    They have been

DR   PDB; 1MOU; X-ray; A=3-221.
DR   PDB; 1MOV; X-ray; A=3-221.
DR   GO; GO:0010117; P:photoprotection; TAS.
DR   GO; GO:0018995; C:host; TAS.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   ProDom; PD013756; Green_fl_protein; 1.
DR   3D-structure; Chromophore; Luminescence; Photoprotein.
KW   3D-structure; Chromophore; Luminescence; Photoprotein.
FT   MOD_RES   63   63   (E)-2,3-didehydrotyrosine.
FT   CROSSLNK   62   64   2-iminomethyl-5-imidazolinone (Gln-Gly).
FT   MUTAGEN   142   142   H->S: Produces a fluorescent form.
FT   MUTAGEN   158   158   F->H: Produces a homodimeric form.
SQ   SEQUENCE   221 AA;   6268IA7BEDD7393B CRC64;

```
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   227 AA;  25640 MW;  56863906F44B244 CRC64;

Query Match             61.4%;  Score 745;  DB 2;  Length 227;
Best Local Similarity   60.8%;  Pred. No. 3.5e-56;
Matches  135;  Conservative  33;  Mismatches  54;  Indels   0;  Gaps   0;

Qy     1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||:||   |||:|||| :| ||||||||:||||| : ||||||||||||||||||||||||
Db     1 MSWSKSVIKEEMLINLHLEGTFNGHYFEIKGKGKGNPNEGTNTVVTLEVTKGGPLPFGWHI 60

Qy    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |:||||||   | ||||||:|||||||||||  |||| | |||:|| | | | || |||||
Db    61 LCPQFQYGNKAFVHHPDDIPDYLKLSFPEGYTWERSMHFEDGGLCCITNDISLTGNCFYY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           | |||| ||| |||||||||:|||| |||||||||||| || || |||| |||:| | |
Db   121 DIKFTGLNFPPNGPVVQKKTTGWEPSTERLYPRDGVLIGDIHHALTVAGGHYVCDIKTV 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHH 222
           | |||  | ||| |||: ||:||:||||||:||:|:|| || |
Db   181 YRAKKPVKMPGYHYVDPKLVIRSNDKEFMKVEEHEIAVARKH 222
```

```
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
-----------------------------------------------------------------------------
DR   EMBL; AF168420; AAF03370.1; -; mRNA.
DR   HSSP; Q9U6Y8; 1GGX.
DR   GO; GO:0008218; P:bioluminescence; ISS.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
KW   Chromophore; Luminescence; Photoprotein.
FT   MOD_RES      66     67       2,3-didehydrotyrosine (By similarity).
FT                67     68       2-iminomethyl-5-imidazolinone (Gln-Gly)
FT                                (By similarity).
FT   CROSSLNK     66     68
SQ   SEQUENCE   232 AA;  26435 MW;  AA8F1BEEE283CE4D CRC64;

Query Match             60.1%;  Score 729.5;  DB 1;  Length 232;
Best Local Similarity   59.1%;  Pred. No. 7.8e-55;
Matches  133;  Conservative  38;  Mismatches  53;  Indels   1;  Gaps   1;

Qy     1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||:|||   |||:|| : ||||||||:||||| : ||||||||||||||||||||||||
Db     1 MSCSKSVIKEEMLIDLHLEGTFNGHYFEIKGKGKGQPNEGTNTVVTLVEVTKGGPLPFGWHI 60

Qy    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |:|||||||   | ||||:|||||||||||  |||| | |||:|| | | | || |||||
Db    61 LCPQFQYGNKAFVHHPDNIHDYLKLSFPEGYTWERSMHFEDGGLCCITNDISLTGNCFYY 120

Qy   121 DIKFTGLNFPPNGPVVQKKTTGWEPSTERLYPRDGVLIGDIHHALTVAGGHYVACDIKTV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   121 DIKFTGLNFPPNGPVVQKKTTGWEPSTERLYPRDGVLIGDIHHALTVAGGHYVACDIKTV 180

Qy   181 YMAKK-PVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLF 224
           | ||   | :||| |||: ||:|: ||||:::|::|:|| || | |
Db   181 YRAKKAALKMPGYHYVDTKLVIWNNDKEFMKVEEHEIAVARHPP 225
```

```
RESULT 15
GPPL_DISST
ID   GPPL_DISST    STANDARD;   PRT;   232 AA.
AC   Q9U6Y7;
DT   01-FEB-2005 (Rel. 46, Created)
DT   01-FEB-2005 (Rel. 46, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   GFP-like fluorescent chromoprotein dsFP483.
OS   Discosoma striata (Striped mushroom).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=105400;
RN   [1]
RP   NUCLEOTIDE SEQUENCE, FUNCTION, AND TISSUE SPECIFICITY.
RX   MEDLINE=99436614; PubMed=10504696; DOI=10.1038/13657;
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RT   "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL   Nat. Biotechnol. 17:969-973(1999).
CC   -!- FUNCTION: Pigment protein that is green in color.
CC   -!- BIOPHYSICOCHEMICAL PROPERTIES:
CC       Absorption:
CC         Abs(max)=443 nm;
CC       Note=Has a strong fluorescence emission spectrum which peaks at
CC       483 nm;
CC   -!- TISSUE SPECIFICITY: Oral disk.
CC   -!- PTM: Contains a chromophore consisting of modified amino acid
CC       residues. The chromophore is formed by autocatalytic backbone
CC       condensation between Xaa-N and Gly-(N+2), oxidation of Tyr-(N+1)
CC       to didehydrotyrosine, and formation of a double bond to the alpha-
CC       amino nitrogen of residue Xaa-N. Maturation of the chromophore
CC       requires nothing other than molecular oxygen. The precise
CC       stereochemistry of the tyrosine has not been determined.
CC   -!- BIOTECHNOLOGY: Fluorescent proteins have become a useful and
CC       ubiquitous tool for making chimeric proteins, where they function
CC       as a fluorescent protein tag. Typically they tolerate N- and C-
CC       terminal fusion to a broad variety of proteins. They have been
CC       expressed in most known cell types and are used as a noninvasive
CC       fluorescent marker in living cells and organisms. They enable a
CC       wide range of applications where they have functioned as a cell
CC       lineage tracer, reporter of gene expression, or as a measure of
CC       protein-protein interactions.
CC   -!- SIMILARITY: Belongs to the GFP family.
```

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 11, 2006, 01:59:57 ; Search time 23 Seconds
808.784 Million cell updates/sec

Title: 1214
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..........EDYIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	225	2	US-09-866-538-12	Sequence 12, Appl
2	1214	100.0	225	2	US-09-865-291-12	Sequence 12, Appl
3	1210	99.7	240	2	US-10-152-296-2	Sequence 2, Appli
4	681	56.1	225	2	US-10-244-779-2	Sequence 2, Appli
5	485.5	40.0	238	2	US-09-277-716-16	Sequence 16, Appl
6	485.5	40.0	238	2	US-09-609-161B-16	Sequence 16, Appl
7	485.5	40.0	238	2	US-09-626-581D-65	Sequence 65, Appl
8	485.5	40.0	238	2	US-09-415-765B-65	Sequence 65, Appl
9	485.5	40.0	238	2	US-09-626-580C-65	Sequence 65, Appl
10	473.5	39.0	224	2	US-09-977-897-3	Sequence 3, Appli
11	473.5	39.0	225	2	US-09-977-897-13	Sequence 13, Appl
12	473.5	39.0	225	2	US-09-977-897-14	Sequence 14, Appl
13	473.5	39.0	226	2	US-09-977-897-15	Sequence 15, Appl
14	473.5	39.0	227	2	US-09-977-897-16	Sequence 16, Appl
15	473.5	39.0	228	2	US-09-977-897-17	Sequence 17, Appl
16	473.5	39.0	229	2	US-09-977-897-18	Sequence 18, Appl
17	473.5	39.0	230	2	US-09-977-897-19	Sequence 19, Appl
18	473.5	39.0	231	2	US-09-977-897-20	Sequence 20, Appl
19	473.5	39.0	232	2	US-09-977-897-21	Sequence 21, Appl
20	473.5	39.0	233	2	US-09-977-897-22	Sequence 22, Appl
21	473.5	39.0	234	2	US-09-977-897-23	Sequence 23, Appl
22	473.5	39.0	234	2	US-09-977-897-9	Sequence 9, Appli
23	473.5	39.0	235	2	US-09-977-897-24	Sequence 24, Appl
24	473.5	39.0	235	2	US-09-977-897-8	Sequence 8, Appli
25	473.5	39.0	236	2	US-09-977-897-7	Sequence 7, Appli
26	473.5	39.0	236	2	US-09-977-897-25	Sequence 25, Appl
27	473.5	39.0	237	2	US-09-977-897-6	Sequence 6, Appli
28	473.5	39.0	237	2	US-09-977-897-26	Sequence 26, Appl
29	473.5	39.0	238	2	US-09-277-716-32	Sequence 32, Appl
30	473.5	39.0	238	2	US-09-609-161B-32	Sequence 32, Appl
31	473.5	39.0	238	2	US-09-839-650-3	Sequence 3, Appli
32	473.5	39.0	238	2	US-09-977-897-5	Sequence 5, Appli
33	473.5	39.0	238	2	US-09-977-897-27	Sequence 27, Appl
34	473.5	39.0	239	2	US-10-652-703A-3	Sequence 3, Appli
35	473.5	39.0	239	2	US-09-977-897-4	Sequence 4, Appli
36	473	39.0	239	2	US-09-977-897-2	Sequence 2, Appli
37	473	39.0	231	2	US-09-977-897-12	Sequence 12, Appl
38	473	39.0	232	2	US-09-977-897-11	Sequence 11, Appl
39	465.5	38.3	233	2	US-09-977-897-10	Sequence 10, Appl
40	465.5	38.3	239	2	US-10-021-818A-4	Sequence 4, Appli
41	215.5	17.8	238	2	US-10-021-818A-23	Sequence 23, Appl
42	214.5	17.7	238	2	US-09-023-946B-23	Sequence 23, Appl
43	214.5	17.7	238	2	US-09-023-946B-28	Sequence 28, Appl
44	213.5	17.6	240	2	US-09-129-192C-49	Sequence 49, Appl
45	212.5	17.5	238	1	US-08-818-604-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-866-538-12
Sequence 12, Application US/09866538
; Patent No. 6852849
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-866-538-12

  Query Match   100.0%;
  Best Local Similarity   100.0%;
  Matches 225;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

  Score 1214;   DB 2;   Length 225;
  Pred. No. 3e-132;

```
Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 2
US-09-865-291-12
Sequence 12, Application US/09865291
; Patent No. 6900304
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin

TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.0

; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-865-291-12

Query Match   100.0%;   Score 1214;   DB 2;   Length 225;
Best Local Similarity   100.0%;
Matches 225;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1   MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120

Qy   121   KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121   KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180

Qy   181   YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181   YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   225

RESULT 3
US-10-152-296-2
; Sequence 2, Application US/10152296
; Patent No. 6723537
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-00001US
; CURRENT APPLICATION NUMBER: US/10/152,296
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian
; OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
; OTHER INFORMATION: "red" red fluorescent protein (RFP)
US-10-152-296-2

Query Match   99.7%;   Score 1210;   DB 2;   Length 240;
Best Local Similarity   99.6%;   Pred. No. 9.5e-132;
Matches 224;   Conservative   1;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1   MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2   VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   61

Qy   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   121

Qy   121   KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122   KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   181

Qy   181   YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   182   YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   226

RESULT 4
US-10-244-779-2
; Sequence 2, Application US/10244779
; Patent No. 6933375
; GENERAL INFORMATION:
; APPLICANT: Falkowski, Paul
; APPLICANT: Sun, Yi
; APPLICANT: Gorbunov, Maxim
; APPLICANT: Wyman, Kevin
; APPLICANT: Chen, Yi-Bu
; TITLE OF INVENTION: mcRP Encoding Nucleic Acids,
; TITLE OF INVENTION: Polypeptides, Antibodies and Methods of Use Thereof
; FILE REFERENCE: Rut 00-0023US
; CURRENT APPLICATION NUMBER: US/10/244,779
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: 60/322,189
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Montastrea cavernosa
US-10-244-779-2

Query Match   56.1%;   Score 681;   DB 2;   Length 225;
Best Local Similarity   57.2%;   Pred. No. 1.7e-70;
Matches 123;   Conservative   36;   Mismatches   56;   Indels   0;   Gaps   0;

Qy   6   NVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF   65
            :||| ||| ||:|:|||:|.:|||||||||.||| |:|::||:.|||||.|| ||:|:|
Db   2   SVIKSVMKIKLRMDGIVNGHKFMITGSEGEGKPFEGTHTIILKVKEGGPLPFAYDILTTAF   61

Qy   66   QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI   125
            ||:.|:||||||:| |||:|:||||.|:|||::|:||.||::|.|:||||:| :.|:||
Db   62   QYGNRVFTKYPKDIPDYFKQSFPEGYSWERSMTFEDGQGVCTVTSDIKLLBGDCFFYEIRFY   121

Qy   126   GVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKK   185
            |||:|:||||||:|:|:||||.||.||||:||||||||:|:|||||||||||.||||.|
Db   122   GVNFPSSGPVMQKKTLKWEPSTENMVRDGVLLGDVSRTLLLBGDKHHRCNFRSTYGAKK   181

Qy   186   PVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGR   220
            :||| ||:|:|| ::|:|||.:|:|||.||:.||
Db   182   GVVLPEYHFVDHRIEILSHDKDYNTVEVYENAVAR   216

RESULT 5
US-09-277-716-16
; Sequence 16, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LU
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238

```
;  TYPE: PRT
;  ORGANISM: Renilla mulleri
;  FEATURE:
;  OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-277-716-16

Query Match     40.0%;  Score 485.5;  DB 2;  Length 238;
Best Local Similarity  44.5%;  Pred. No. 8.5e-48;
Matches 97;  Conservative 43;  Mismatches 71;  Indels 7;  Gaps 3;

Qy      4 SKNVIK------EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW 58
            |||| ::|       |:|:::::|| ||:| |||:|:||||| | ||| ||||||:
Db      2 SKQILKNTCLQEVMSYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF 61

Qy     59 DILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCF 118
            |||:||||||||||:|||:| |:| |||||||||||:|||||||||||||||:|||||:
Db     62 DIVSPAFQYGNRTFYKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF 121

Qy    119 IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFK 178
            | |:|| ||||| ||| |||||| ||:|  |||||| |||||:|||:| |||||||| :|
Db    122 VYRVEYKGSNFPDDGPVMQKTILGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK 181

Qy    179 SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
            ::| : :| ||||||| |:|||:|||:|||||:|||:|
Db    182 TLMKSKGVVKEPPSYHFIQHRLEKT-YVEDGGFVEQHE 218


RESULT 6
US-09-609-161B-16
; Sequence 16, Application US/0960916lB
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; TITLE OF INVENTION: PROLUME, LTD
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-09-609-161B-16

Query Match     40.0%;  Score 485.5;  DB 2;  Length 238;
Best Local Similarity  44.5%;  Pred. No. 8.5e-48;
Matches 97;  Conservative 43;  Mismatches 71;  Indels 7;  Gaps 3;

Qy      4 SKNVIK------EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW 58
            |||| ::|       |:|:::::|| ||:| |||:|:||||| | ||| ||||||:
Db      2 SKQILKNTCLQEVMSYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF 61

Qy     59 DILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCF 118
            |||:||||||||||:|||:| |:| |||||||||||:|||||||||||||||:|||||:
Db     62 DIVSPAFQYGNRTFYKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF 121

Qy    119 IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFK 178
            | |:|| ||||| ||| |||||| ||:|  |||||| |||||:|||:| |||||||| :|
Db    122 VYRVEYKGSNFPDDGPVMQKTILGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK 181

Qy    179 SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
            ::| : :| ||||||| |:|||:|||:|||||:|||:|
Db    182 TLMKSKGVVKEPPSYHFIQHRLEKT-YVEDGGFVEQHE 218


RESULT 7
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-581D-65

Query Match     40.0%;  Score 485.5;  DB 2;  Length 238;
Best Local Similarity  44.5%;  Pred. No. 8.5e-48;
Matches 97;  Conservative 43;  Mismatches 71;  Indels 7;  Gaps 3;

Qy      4 SKNVIK------EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW 58
            |||| ::|       |:|:::::|| ||:| |||:|:||||| | ||| ||||||:
Db      2 SKQILKNTCLQEVMSYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF 61

Qy     59 DILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCF 118
            |||:||||||||||:|||:| |:| |||||||||||:|||||||||||||||:|||||:
Db     62 DIVSPAFQYGNRTFYKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF 121

Qy    119 IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFK 178
            | |:|| ||||| ||| |||||| ||:|  |||||| |||||:|||:| |||||||| :|
Db    122 VYRVEYKGSNFPDDGPVMQKTILGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK 181

Qy    179 SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
            ::| : :| ||||||| |:|||:|||:|||||:|||:|
Db    182 TLMKSKGVVKEPPSYHFIQHRLEKT-YVEDGGFVEQHE 218


RESULT 8
US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-415-765B-65

Query Match     40.0%;  Score 485.5;  DB 2;  Length 238;
```

RESULT 9

US-09-626-580C-65
Sequence 65, Application US/0962658OC
Patent No. 6562617

; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-580C-65

```
  Query Match      40.0%;   Score 485.5;   DB 2;   Length 238;
  Best Local Similarity   44.5%;   Pred. No. 8.5e-48;
  Matches 97;   Conservative 43;   Mismatches 71;   Indels 7;   Gaps 3;

Qy    4  SKNVIK------EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW   58
           :  ::|      |::|| :|   | |:|  ||:||:||:|| :|  ||| ||| ||||| |
Db    2  SKQILKNTCLQEVMSVYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF   61

Qy   59  DILSPQFQYGSKVYVKHPADIIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCF  118
           ||:| |||||||| ||||| |:||  |||||||||||||| | |:|||||||:|| || |
Db   62  DIVSPAFQYGNRTFTKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF   121

Qy  119  IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFK  178
           ||: |||||||||||||:|||||||||||||||| :|||||||  |||  |||||:||||
Db  122  VYRVEYKGSNFPDDGPVMQKTILGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK   181

Qy  179  SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE   215
           |||| |||| :|||||||||||| :| |||||| |:|
Db  182  TLMKSKGVVKEFPSYHFIQHRLEKT-YVEDGGFVEQHE   218
```

RESULT 10

US-09-977-897-3
Sequence 3, Application US/0997897
Patent No. 6780974

; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce

; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-3

```
  Query Match      39.0%;   Score 473.5;   DB 2;   Length 224;
  Best Local Similarity   45.5%;   Pred. No. 1.9e-46;
  Matches 97;   Conservative 40;   Mismatches 73;   Indels 3;   Gaps 3;

Qy    5  KNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP   63
          ||  |::|| :| | |:|  ||:||:|| ||:|| :|  ||| ||||||||||||  :|
Db    8  KNTGLKEIMSAKASVEGIVNNHVFSMEGFGKGNVLFGNQLMQIRVTKGGPLPFAFDIVSI   67

Qy   64  QFQYGSKVYVKHPADIIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVK  123
          :||||| | ||||| |:||  |||||| | ||| |:|||| : |:| |:|::| ||:||
Db   68  AFQYGNRTFTKYPDDIADYFVQSFPAGFFYERNLRFEDGAIVDIRSDISLEDDKFHVKVE   127

Qy  124  FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA  183
          :  |||| |||||||| || :|| || | :|||||| |||||  ||:||||  ||:| |
Db  128  YRGNGFPSNGPVMQKAILGMEPSFEVVYMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS   187

Qy  184  KKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE   215
          |  |  :| |||||||| :| |||||| |:|
Db  188  KGGVKEFPEYHFIHHRLEKT-YVEEGSFVEQHE   219
```

RESULT 11

US-09-977-897-13
Sequence 13, Application US/0997897
Patent No. 6780974

; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescen
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems and
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-13

```
  Query Match      39.0%;   Score 473.5;   DB 2;   Length 224;
  Best Local Similarity   45.5%;   Pred. No. 1.9e-46;
  Matches 97;   Conservative 40;   Mismatches 73;   Indels 3;   Gaps 3;

Qy    4  SKNVIK------EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW   58
           :  ::|      |::|| :|   | |:|  ||:||:||:|| :|  ||| ||| ||||| |
Db    2  SKQILKNTCLQEVMSVYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF   61

Qy   59  DILSPQFQYGSKVYVKHPADIIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCF  118
           ||:| |||||||| ||||| |:||  |||||||||||||| | |:|||||||:|| || |
Db   62  DIVSPAFQYGNRTFTKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF   121

Qy  119  IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFK  178
           ||: |||||||||||||:|||||||||||||||| :|||||||  |||  |||||:||||
Db  122  VYREYKGSNFPDDGPVMQKTILGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK   181

Qy  179  SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE   215
           |||| |||| :|||||||||||| :| |||||| |:|
Db  182  TLMKSKGVVKEFPSYHFIQHRLEKT-YVEDGGFVEQHE   218
```

RESULT 12
US-09-977-897-14
; Sequence 14, Application US/0977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 14
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-14

Query Match   39.0%; Score 473.5; DB 2; Length 225;
Best Local Similarity   45.5%; Pred. No. 1.9e-46;
Matches   97; Conservative   40; Mismatches   73; Indels   3; Gaps   3;

Qy   5 KNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP 63
           : ::::::::::::::::::::::::::: ::::::::::::::::::::::::::::: :::
Db   8 KNTGLKEIMSAKASVEGIVNNHVFSMEGFGKGNVLFGNQLMQIRVTKGGPLPFAFDIVSI 67

Qy   64 QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIYKVK 123
           :::::: :::::::: ::: :::::::: :::: :::::::::::::::::::::::: :::
Db   68 YRGNGPPSNGPVWQKAILGMEPSFEVVYMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS 127

Qy   124 FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
           :::
Db   128 YRGNGPPSNGPVWQKAILGMEPSFEVVYMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy   184 KKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
           :::: :::::::::::::::::::::::::::
Db   188 KGGVKEFPEYHFIHHRLEKT-YVEEGSFVEQHE 219

RESULT 13
US-09-977-897-15
; Sequence 15, Application US/0977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 15
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-15

Query Match   39.0%; Score 473.5; DB 2; Length 226;
Best Local Similarity   45.5%; Pred. No. 1.9e-46;
Matches   97; Conservative   40; Mismatches   73; Indels   3; Gaps   3;

Qy   5 KNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP 63
           : ::::::::::::::::::::::::::: ::::::::::::::::::::::::::::: :::
Db   8 KNTGLKEIMSAKASVEGIVNNHVFSMEGFGKGNVLFGNQLMQIRVTKGGPLPFAFDIVSI 67

Qy   64 QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIYKVK 123
           :::::: :::::::: ::: :::::::: :::: :::::::::::::::::::::::: :::
Db   68 AFQYGNRTFTKYDDIADYFVQSFPAGFFYERNLRFEDGAIVDIRSDISLEDDKFHYKVE 127

Qy   124 FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
           :::
Db   128 YRGNGPPSNGPVWQKAILGMEPSFEVVYMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy   184 KKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
           :::: :::::::::::::::::::::::::::
Db   188 KGGVKEFPEYHFIHHRLEKT-YVEEGSFVEQHE 219

RESULT 14
US-09-977-897-16
; Sequence 16, Application US/0977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems and
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 16
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-16

Query Match   39.0%; Score 473.5; DB 2; Length 227;
Best Local Similarity   45.5%; Pred. No. 1.9e-46;
Matches   97; Conservative   40; Mismatches   73; Indels   3; Gaps   3;

Qy   5 KNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP 63
           : ::::::::::::::::::::::::::: ::::::::::::::::::::::::::::: :::
Db   8 KNTGLKEIMSAKASVEGIVNNHVFSMEGFGKGNVLFGNQLMQIRVTKGGPLPFAFDIVSI 67

Qy   64 QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIYKVK 123
           :::::: :::::::: ::: :::::::: :::: :::::::::::::::::::::::: :::
Db   68 AFQYGNRTFTKYDDIADYFVQSFPAGFFYERNLRFEDGAIVDIRSDISLEDDKFHYKVE 127

Qy   124 FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
           :::
Db   128 YRGNGPPSNGPVWQKAILGMEPSFEVVYMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy   184 KKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
           :::: :::::::::::::::::::::::::::
Db   188 KGGVKEFPEYHFIHHRLEKT-YVEEGSFVEQHE 219

RESULT 15
US-09-977-897-17
; Sequence 17, Application US/0977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems and
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 17
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-17

Query Match   39.0%; Score 473.5; DB 2; Length 228;
Best Local Similarity   45.5%; Pred. No. 2e-46;

Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

```
Qy     5 KQNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTIKGGPLPFAWDILSP  63
           ||  :||| |   ::||  || |: :::|:|||||||||:|:|
Db     8 KNTGLKEIMSAKASVEGIVNNHVFSMEGRGKGNVLFGNQLMQIRVTIKGGPLPFAFDIVSI  67

Qy    64 QFQYGSKVYVKHPADIIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLLQDGCFIYKVK 123
          |||| || :| ||| :|:| : |||:||| |||  :|||| :| | |:|:|||| |:|||:
Db    68 AFQYGNRTFTKYPDDIADYFVQSFPAGFYYERNLRFEDGAIVDIRSDISLEDDKFHYKVE  127

Qy   124 PIGVNFPSDGPVMQKKTMGWEASTERLIYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
          : || ||| |||||| :||  |||:|  : ||| || | || :| |||:| |:  ::||
Db   128 YRGNGFPSNGPVMQKAILGMEPSFEVVYMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS  187

Qy   184 KKPV-QLPGYYYVVDSKLDITSHNEDYTIVEQYE 215
          |  | :| :|:: |: :: |:: |||:|:|||:|
Db   188 KGGVKEPPEYHPIHHRLEKT-YVEEGSFVEQHE  219
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2006 Compugen Ltd.

GenCore version 5.1.6

Run on: January 11, 2006, 02:00:37 ; Search time 62 Seconds
                                           (without alignments)
                                           1516.316 Million cell updates/sec

Title: 1214
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published_Applications_AA_Main:*
                 1: /cgn2_6/ptcdata/1/pubpaa/US07_PUBCOMB.pep:*
                 2: /cgn2_6/ptcdata/1/pubpaa/US08_PUBCOMB.pep:*
                 3: /cgn2_6/ptcdata/1/pubpaa/US09_PUBCOMB.pep:*
                 4: /cgn2_6/ptcdata/1/pubpaa/US10A_PUBCOMB.pep:*
                 5: /cgn2_6/ptcdata/1/pubpaa/US10B_PUBCOMB.pep:*
                 6: /cgn2_6/ptcdata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match DB Length	ID	Description
1	1214	100.0	225	3	US-09-999-745-67	Sequence 67, Appli
2	1214	100.0	225	3	US-09-866-538-12	Sequence 12, Appl
3	1214	100.0	225	3	US-09-794-308-12	Sequence 12, Appl
4	1214	100.0	225	3	US-09-865-291-12	Sequence 12, Appl
5	1214	100.0	225	4	US-10-006-922-12	Sequence 12, Appl
6	1214	100.0	225	4	US-10-081-864-8	Sequence 8, Appli
7	1214	100.0	225	4	US-10-121-258-1	Sequence 1, Appli
8	1214	100.0	225	4	US-10-315-920-2	Sequence 2, Appli
9	1214	100.0	225	4	US-10-132-067-4	Sequence 4, Appl
10	1214	100.0	225	4	US-10-370-570-56	Sequence 56, Appl
11	1214	100.0	225	4	US-10-406-618-32	Sequence 32, Appl
12	1214	100.0	225	4	US-10-433-640-13	Sequence 13, Appl
13	1214	100.0	225	4	US-10-311-030-7	Sequence 7, Appli
14	1214	100.0	225	5	US-10-885-484-4	Sequence 3, Appli
15	1214	100.0	225	5	US-10-885-988-12	Sequence 12, Appl
16	1214	100.0	225	5	US-10-656-029-222	Sequence 22, Appl
17	1214	100.0	225	5	US-10-857-622-12	Sequence 12, Appl
18	1214	100.0	225	5	US-10-505-486-26	Sequence 26, Appl
19	1214	100.0	225	5	US-10-844-064A-2	Sequence 2, Appl
20	1214	100.0	225	5	US-10-931-304-1	Sequence 1, Appl
21	1214	100.0	487	5	US-10-343-977-1	Sequence 1, Appli
22	1214	100.0	487	5	US-10-343-977-1	Sequence 1, Appli
23	1214	100.0	506	5	US-10-343-977-2	Sequence 2, Appli
24	1214	100.0	506	5	US-10-343-977-2	Sequence 2, Appli
25	1214	100.0	547	4	US-10-343-977-3	Sequence 3, Appli
26	1214	100.0	547	4	US-10-343-977-3	Sequence 3, Appli
27	1210	99.7	226	4	US-10-311-030-9	Sequence 9, Appli
28	1210	99.7	240	4	US-10-152-296-2	Sequence 2, Appli
29	1210	99.7	240	4	US-10-739-656-2	Sequence 2, Appl
30	1210	99.7	545	4	US-10-214-932-52	Sequence 52, Appl
31	1210	99.7	548	4	US-10-214-932-76	Sequence 76, Appl
32	1207	99.4	225	4	US-10-121-258-20	Sequence 20, Appl
33	1207	99.4	225	5	US-10-315-920-4	Sequence 4, Appli
34	1207	99.4	225	5	US-10-931-304-20	Sequence 20, Appli
35	1196	98.5	236	4	US-10-314-936-4	Sequence 4, Appli
36	1196	98.5	236	6	US-10-314-936-4	Sequence 4, Appli
37	1196	98.5	236	4	US-11-021-014-2	Sequence 2, Appli
38	1196	98.5	236	4	US-11-021-014-4	Sequence 4, Appli
39	1196	98.5	236	6	US-10-724-178-12	Sequence 12, Appl
40	1195.5	98.5	225	4	US-10-315-920-6	Sequence 6, Appli
41	1191	98.1	225	4	US-10-442-148A-7	Sequence 7, Appli
42	1186	97.7	239	4	US-10-442-148A-8	Sequence 8, Appli
43	1186	97.7	225	4	US-10-121-258-4	Sequence 4, Appli
44	1160	95.6	225	5	US-10-931-304-4	Sequence 4, Appli
45	1153	95.0	225	4	US-10-121-258-24	Sequence 24, Appli

ALIGNMENTS

RESULT 1
US-09-999-745-67

; Sequence 67, Application US/0999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.

US-09-999-745-67

Query Match   100.0%; Score 1214; DB 3; Length 225;
Best Local Similarity   100.0%;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 2
US-09-866-538-12

; Sequence 12, Application US/09865538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger

```
;       APPLICANT: Campbell, Robert
;       TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
;       FILE REFERENCE: REGEN1530-2
;       CURRENT APPLICATION NUMBER:
;       CURRENT FILING DATE: 2001-05-24
;       NUMBER OF SEQ ID NOS: 29
;       SOFTWARE: Patentin version 3.0
;
;       SEQ ID NO 12
;       LENGTH: 225
;       TYPE: PRT
;       ORGANISM: Discosoma sp.
US-09-866-538-12

        Query Match             100.0%;  Score 1214;  DB 3;  Length 225;
        Best Local Similarity   100.0%;  Pred. No. 6.6e-114;
        Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy       61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy      121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 3
US-09-794-308-12
;  Sequence 12, Application US/09794308
;  Publication No. US20030170911A1
;  GENERAL INFORMATION:
;       APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
;       APPLICANT: TSIEN, Roger
;       APPLICANT: ZACHARIAS, David
;       APPLICANT: BAIRD, Geoffrey
;       TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
;       FILE REFERENCE: REGEN1530
;       CURRENT APPLICATION NUMBER: US/09/794,308
;       CURRENT FILING DATE: 2001-02-26
;       NUMBER OF SEQ ID NOS: 25
;       SOFTWARE: Patentin version 3.0
;
;       SEQ ID NO 12
;       LENGTH: 225
;       TYPE: PRT
;       ORGANISM: Discosoma sp.
US-09-794-308-12

        Query Match             100.0%;  Score 1214;  DB 3;  Length 225;
        Best Local Similarity   100.0%;  Pred. No. 6.6e-114;
        Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy       61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy      121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
Db      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 4
US-09-865-291-12
;  Sequence 12, Application US/09865291
;  Publication No. US20030186229A1
;  GENERAL INFORMATION:
;       APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
;       APPLICANT: TSIEN, Roger
;       APPLICANT: TING, Alice
;       APPLICANT: ZHANG, Jin
;       TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
;       FILE REFERENCE: REGEN1550
;       CURRENT APPLICATION NUMBER: US/09/865,291
;       CURRENT FILING DATE: 2001-05-24
;       NUMBER OF SEQ ID NOS: 42
;       SOFTWARE: Patentin version 3.0
;
;       SEQ ID NO 12
;       LENGTH: 225
;       TYPE: PRT
;       ORGANISM: Discosoma sp.
US-09-865-291-12

        Query Match             100.0%;  Score 1214;  DB 3;  Length 225;
        Best Local Similarity   100.0%;  Pred. No. 6.6e-114;
        Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy       61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy      121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db      181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 5
US-10-006-922-12
;  Sequence 12, Application US/10006922
;  Publication No. US20020197676A1
;  GENERAL INFORMATION:
;       APPLICANT: Lukyanov, Sergey A
;       APPLICANT: Fradkov, Arcady F.
;       APPLICANT: Labas, Yulii A.
;       APPLICANT: Matz, Mikhail V.
;       APPLICANT: Terskikh, Alexey
;       TITLE OF INVENTION: No. US20020197676A1e1 Chromophores/Fluorophores and
;       TITLE OF INVENTION: Methods for Using the Same
;       FILE REFERENCE: CLON-035CIP
;       CURRENT APPLICATION NUMBER: US/10/006,922
;       CURRENT FILING DATE: 2001-12-04
;       PRIOR APPLICATION NUMBER: 09/120,330
;       PRIOR FILING DATE: 1998-12-11
;       PRIOR APPLICATION NUMBER: 09/457,898
;       PRIOR FILING DATE: 1999-12-09
;       PRIOR APPLICATION NUMBER: 09/458,144
;       PRIOR FILING DATE: 1999-12-09
;       PRIOR APPLICATION NUMBER: 09/458,477
;       PRIOR FILING DATE: 1999-12-09
;       PRIOR APPLICATION NUMBER: 09/457,556
;       PRIOR FILING DATE: 1999-12-09
;       PRIOR APPLICATION NUMBER: 09/444,338
;       PRIOR FILING DATE: 1999-11-19
;       NUMBER OF SEQ ID NOS: 46
```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-12

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 6
US-10-081-864-8
; Sequence 8, Application No. US10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 7
US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Talen, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(225)
; OTHER INFORMATION: wild-type DsRed
US-10-121-258-1

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 8
US-10-315-920-2
; Sequence 2, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT

; ORGANISM: Discosoma sp.
US-10-315-920-2

Query Match   100.0%;   Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%;   Pred. No. 6.6e-114;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 9
US-10-132-067-4

; Sequence 4, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorescence
; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
; TITLE OF INVENTION: Fluorescence
; FILE REFERENCE: 021362-00060US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; OTHER INFORMATION: red fluorescent protein (dsRED)
US-10-132-067-4

Query Match   100.0%;   Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%;   Pred. No. 6.6e-114;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 10
US-10-370-570-56

; Sequence 56, Application US/10370570
; Publication No. US20030219717A1

; GENERAL INFORMATION:
; APPLICANT: DAHL, Soren Weis et al.
; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS
; FILE REFERENCE: 3759-0130P
; CURRENT APPLICATION NUMBER: US/10/370,570
; CURRENT FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 56
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-370-570-56

Query Match   100.0%;   Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%;   Pred. No. 6.6e-114;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 11
US-10-406-618-32

; Sequence 32, Application US/10406618
; Publication No. US20030219814A1
; GENERAL INFORMATION:
; APPLICANT: Wan, David Chi-Cheong
; APPLICANT: Ip, Denis Tsz-Ming
; APPLICANT: The Chinese University of Hong Kong
; TITLE OF INVENTION: No. US20030219814Ale1 Fluorescent Proteins
; FILE REFERENCE: 016285-34-1
; CURRENT APPLICATION NUMBER: US/10/406,618
; CURRENT FILING DATE: 2003-04-02
; PRIOR APPLICATION NUMBER: US 60/387,968
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: US 60/370,598
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 32
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Spontaneously
; OTHER INFORMATION: fluorescent protein DsRed.
US-10-406-618-32

Query Match   100.0%;   Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%;   Pred. No. 6.6e-114;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

RESULT 12
US-10-433-640-13
; Sequence 13, Application US/10433640
; Publication No. US20040115792A1
; GENERAL INFORMATION:
; APPLICANT: Lichtenberg-Fraté, Hella
; TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
; TITLE OF INVENTION: COMPLEX ENVIRONMENTAL CONTAMINATIONS
; FILE REFERENCE: 1487/3
; CURRENT APPLICATION NUMBER: US/10/433,640
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/14610
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: DE 10061872.3
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 13
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-433-640-13

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity 100.0%;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 13
US-10-311-030-7
; Sequence 7, Application US/10311030
; Publication No. US20040117110A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamlara, Elize
; APPLICANT: Telen, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/184,732
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-311-030-7

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity 100.0%;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 14
US-10-845-484-3
; Sequence 3, Application US/10845484
; Publication No. US20040248180A1
; GENERAL INFORMATION:
; APPLICANT: Bulina, Maria
; APPLICANT: Chudakov, Dmitry
; APPLICANT: Lukyanov, Konstantin
; TITLE OF INVENTION: Mutant Chromophores/Flourophores and
; TITLE OF INVENTION: Methods for Making and Using the Same
; FILE REFERENCE: CLON 092
; CURRENT APPLICATION NUMBER: CLON 092
; CURRENT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: US/10/845,484
; PRIOR FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: 60/343128
; PRIOR FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: PCT/US02/41418
; PRIOR FILING DATE: 2002-12-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-845-484-3

Query Match   100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 15
US-10-885-988-12
; Sequence 12, Application US/10885988
; Publication No. US20040259165A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/10/885,988
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-885-988-12

Query Match   100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Search completed: January 11, 2006, 02:06:37
Job time : 63 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 11, 2006, 02:03:07 ; Search time 8 Seconds
(without alignments)
238.884 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 61141 seqs, 8493638 residues

Total number of hits satisfying chosen parameters: 61141

Minimum Match 0%
Maximum Match 100%

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	225	6	US-10-209-208-1	Sequence 1, Appl
2	1207	99.4	225	6	US-10-209-208-20	Sequence 20, Appl
3	1160	95.6	225	6	US-10-209-208-4	Sequence 4, Appl
4	1153	95.0	225	6	US-10-209-208-24	Sequence 24, Appl
5	1121	92.3	226	6	US-10-209-208-6	Sequence 6, Appl
6	1021	84.1	225	6	US-10-209-208-8	Sequence 8, Appl
7	1012	83.4	225	6	US-10-209-208-79	Sequence 79, Appl
8	485.5	40.0	238	7	US-11-175-411-16	Sequence 16, Appl
9	485.5	40.0	238	7	US-11-175-766-16	Sequence 16, Appl
10	473.5	39.0	238	7	US-11-179-411-32	Sequence 32, Appl
11	473.5	39.0	238	7	US-11-175-766-32	Sequence 32, Appl
12	422.5	34.8	233	7	US-11-175-411-27	Sequence 27, Appl
13	422.5	34.8	233	7	US-11-175-766-27	Sequence 27, Appl
14	410.5	33.8	233	7	US-11-175-411-33	Sequence 33, Appl
15	410.5	33.8	233	7	US-11-175-766-33	Sequence 33, Appl
16	210.5	17.3	238	6	US-10-209-208-10	Sequence 10, Appl
17	210.5	17.3	239	7	US-11-089-551A-40	Sequence 40, Appl
18	209.5	17.3	239	6	US-10-209-208-12	Sequence 12, Appl
19	205.5	16.9	239	6	US-10-209-208-15	Sequence 15, Appl
20	204.5	16.8	512	7	US-11-032-236-6	Sequence 6, Appl
21	204.5	16.8	1219	7	US-11-032-236-4	Sequence 4, Appl
22	200.5	16.5	239	6	US-10-209-208-13	Sequence 13, Appl
23	192.5	15.9	239	6	US-10-209-208-11	Sequence 11, Appl
24	192.5	15.9	239	6	US-10-209-208-14	Sequence 14, Appl
25	183.5	15.1	437	7	US-11-069-642-92	Sequence 92, Appl
26	182.5	15.0	438	7	US-11-069-642-47	Sequence 47, Appl
27	182.5	15.0	438	7	US-11-069-642-49	Sequence 49, Appl
28	182.5	15.0	438	7	US-11-069-642-51	Sequence 51, Appl
29	182.5	15.0	438	7	US-11-069-642-53	Sequence 53, Appl
30	182.5	15.0	438	7	US-11-069-642-55	Sequence 55, Appl
31	182.5	15.0	438	7	US-11-069-642-57	Sequence 57, Appl
32	182.5	15.0	438	7	US-11-069-642-59	Sequence 59, Appl
33	182.5	15.0	438	7	US-11-069-642-61	Sequence 61, Appl
34	182.5	15.0	438	7	US-11-069-642-63	Sequence 63, Appl
35	103	8.5	559	7	US-10-793-626-1376	Sequence 1376, Ap
36	80	6.6	424	6	US-10-467-657-90	Sequence 90, Appl
37	80	6.6	424	6	US-10-467-657-6344	Sequence 6344, Ap
38	80	6.6	690	7	US-11-052-55A-232	Sequence 232, App
39	76.5	6.3	325	6	US-10-467-657-5568	Sequence 5568, Ap
40	76.5	6.3	574	7	US-10-467-657-14	Sequence 14, Appl
41	76.5	6.3	724	7	US-11-053-100-40	Sequence 40, Appl
42	76.5	6.3	874	7	US-11-053-100-41	Sequence 41, Appl
43	76.5	6.3	1174	7	US-11-053-100-42	Sequence 42, Appl
44	76.5	6.3	1174	6	US-11-053-100-43	Sequence 43, Appl
45	76.5	6.3	1663	6	US-10-055-877-148	Sequence 148, App

ALIGNMENTS

RESULT 1
US-10-209-208-1

Sequence 1, Application US10209208
Publication No. US20050244921A1

GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
APPLICANT: Geoffrey Baird
TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
TITLE OF INVENTION: FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP2
CURRENT APPLICATION NUMBER: US/10/209,208
CURRENT FILING DATE:
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 80
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 1
LENGTH: 225
TYPE: PRT
ORGANISM: Discosoma sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(225)
OTHER INFORMATION: wild-type DsRed

US-10-209-208-1

Query Match 100.0%; Score 1214; DB 6; Length 225;
Best Local Similarity 100.0%; Pred. No. 2.3e-110;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 2
US-10-209-208-20
; Sequence 20, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed with I125R
US-10-209-208-20

Query Match   99.4%; Score 1207; DB 6; Length 225;
Best Local Similarity 99.6%; Pred. No. 1.1e-109;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 3
US-10-209-208-4
; Sequence 4, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1" with I125R mutation
US-10-209-208-4

Query Match   95.0%; Score 1153; DB 6; Length 225;
Best Local Similarity 95.6%; Pred. No. 1.8e-104;
Matches 215; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          | ||:||||||||||||||||| ||||||||||||||||||:||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

```
Db  121  KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
              |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL  225
```

RESULT 5
US-10-209-208-6
Sequence 6, Application US/10209208
Publication No. US20050244921A1
GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  APPLICANT: Campbell, Geoffrey Baird
  TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS FOR MAKING SAME
  FILE REFERENCE: UC083.1CP2CP2
  CURRENT APPLICATION NUMBER: US/10/209,208
  CURRENT FILING DATE: 2002-07-29
  PRIOR APPLICATION NUMBER: 10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  NUMBER OF SEQ ID NOS: 80
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 6
  LENGTH: 226
  TYPE: PRT
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: DsRed polypeptide variant "dimer2"
US-10-209-208-6

Query Match   92.3%;   Score 1121;   DB 6;   Length 226;
Best Local Similarity   92.8%;   Pred. No. 2.2e-101;
Matches 207; Conservative   7; Mismatches   9; Indels   0; Gaps   0;

```
Qy    3  SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS   62
         |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4  SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS   63

Qy   63  PQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKV  122
         |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64  PQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV  123

Qy  123  KFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYM  182
         ||:|:||||||||||||||||||||||||||||||||:|||||||||||||||||||:|||
Db  124  KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYM  183

Qy  183  AKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||:|||||||||||||||||||:|||||||
Db  184  AKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFL  226
```

RESULT 6
US-10-209-208-8
Sequence 8, Application US/10209208
Publication No. US20050244921A1
GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  APPLICANT: Campbell, Geoffrey Baird
  TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS FOR MAKING SAME
  FILE REFERENCE: UC083.1CP2CP2
  CURRENT APPLICATION NUMBER: US/10/209,208
  CURRENT FILING DATE: 2002-07-29
  PRIOR APPLICATION NUMBER: 10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  NUMBER OF SEQ ID NOS: 80
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 8
  LENGTH: 225
  TYPE: PRT
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: DsRed polypeptide variant "mRFP1"
US-10-209-208-8

Query Match   84.1%;   Score 1021;   DB 6;   Length 225;
Best Local Similarity   86.9%;   Pred. No. 1e-91;
Matches 192; Conservative   5; Mismatches 24; Indels   0; Gaps   0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         |||::||||||||||||||||:|||||||||||||||||||||:::||||||||||||||
Db    1  MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||::||
Db   61  LSPQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||:|:|:||||:|||||||||||||||||||||||||||||::|||||||||||||||||
Db  121  KVKLRGTNFPPDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH  221
         :||||||||||||||||:|||||||||||||||||||||||
Db  181  YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH  221
```

RESULT 7
US-10-209-208-79
Sequence 79, Application US/10209208
Publication No. US20050244921A1
GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  APPLICANT: Campbell, Geoffrey Baird
  TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS FOR MAKING SAME
  FILE REFERENCE: UC083.1CP2CP2
  CURRENT APPLICATION NUMBER: US/10/209,208
  CURRENT FILING DATE: 2002-07-29
  PRIOR APPLICATION NUMBER: 10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  NUMBER OF SEQ ID NOS: 80
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 79
  LENGTH: 225
  TYPE: PRT
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: DsRed polypeptide variant "mRFP1.1"
US-10-209-208-79

Query Match   83.4%;   Score 1012;   DB 6;   Length 225;
Best Local Similarity   86.0%;   Pred. No. 7.4e-91;
Matches 190; Conservative   6; Mismatches 25; Indels   0; Gaps   0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         |||::||||||||||||||||:|||||||||||||||||||||:::||||||||||||||
Db    1  MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||::||
Db   61  LSPQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIY  120
```

RESULT 8
US-11-179-411-16
; Sequence 16, Application US/11179411
; Publication No. US20050264991A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/179,411
; CURRENT FILING DATE: 2005-07-12
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 08/757,046
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 09/597,274
; PRIOR FILING DATE: 1996-02-06
; PRIOR APPLICATION NUMBER: 08/908,909
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 08/990,103
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-11-179-411-16

Query Match   40.0%; Score 485.5; DB 7; Length 238;
Best Local Similarity 44.5%; Pred. No. 5.8e-40;
Matches 97; Conservative 43; Mismatches 71; Indels 7; Gaps 3;

```
Qy    4 SKNVIK-----EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW  58
         ||::|     :||:|||||||||||::|::| |:::|||| :::|||||:|||||||
Db    2 SKQILKNTCLQEVMSYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF  61

Qy   59 DILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVLKGEIHKALKLKDGGHYLVEFK 118
         ||:||:||||.:::.|.||:.||::|||| ||:::||||:||||.:||||::.||||:||:||||||
Db   62 DIVSPAFQYGNRTFTKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF 121

Qy  119 IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLIYPRDGVLKGEIHKALKLKDGGHYLVEFK 178
         :|:|:..|.|||.||||||||:|:|.|.|..:.|.:.:.||||::.||||:||:||||||
Db  122 VYRVEYKGSNFPDDGPVMQKITLGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK 181

Qy  179 SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
         :: ||:...|:||:|::||:||||||||||||||||||
Db  182 TLMKSKGVVKEFPSYHFIQHRLEKT-YVEDGGFVEQHE 218
```

RESULT 9
US-11-175-766-16
; Sequence 16, Application US/11175766
; Publication No. US20050272111A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/175,766
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 08/757,046
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 09/597,274
; PRIOR FILING DATE: 1996-02-06
; PRIOR APPLICATION NUMBER: 08/908,909
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 08/990,103
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-11-175-766-16

Query Match   40.0%; Score 485.5; DB 7; Length 238;
Best Local Similarity 44.5%; Pred. No. 5.8e-40;
Matches 97; Conservative 43; Mismatches 71; Indels 7; Gaps 3;

```
Qy    4 SKNVIK-----EFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAW  58
         ||::|     :||:|||||||||||::|::| |:::|||| :::|||||:|||||||
Db    2 SKQILKNTCLQEVMSYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFAF  61

Qy   59 DILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVLKGEIHKALKLKDGGHYLVEFK 118
Db   62 DIVSPAFQYGNRTFTKYPNDISDYFIQSFPAGFMYERTLRYEDGGLVEIRSDINLIEDKF 121

Qy  119 IYKVKFIGVNFPSDGPVMQKKTMGWEASTERLIYPRDGVLKGEIHKALKLKDGGHYLVEFK 178
Db  122 VYRVEYKGSNFPDDGPVMQKITLGIEPSFEAMYMNNGVLVGEVILVYKLNSGKYYSCHMK 181

Qy  179 SIYMAKKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
Db  182 TLMKSKGVVKEFPSYHFIQHRLEKT-YVEDGGFVEQHE 218
```

RESULT 10
US-11-179-411-32
; Sequence 32, Application US/11179411
; Publication No. US20050264991A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/179,411
; CURRENT FILING DATE: 2005-07-12
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26

```
;   PRIOR APPLICATION NUMBER: 08/757,046
;   PRIOR FILING DATE: 1996-11-25
;   PRIOR APPLICATION NUMBER: 08/597,274
;   PRIOR FILING DATE: 1996-02-06
;   PRIOR APPLICATION NUMBER: 08/908,909
;   PRIOR FILING DATE: 1997-08-08
;   PRIOR APPLICATION NUMBER: 08/990,103
;   PRIOR FILING DATE: 1997-12-12
;   NUMBER OF SEQ ID NOS: 33
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 32
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Ptilosarcus gurneyi
US-11-179-411-32

  Query Match       39.0%; Score 473.5; DB 7; Length 238;
  Best Local Similarity   45.5%; Pred. No. 8.3e-39;
  Matches   97; Conservative   40; Mismatches   73; Indels   3; Gaps   3;

Qy    5 KNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP  63
        ||| :|| |::|| : | | ||||:|||| ::||||||||||| | |||||||||:||:|:
Db    7 KNTGLKEIMSAKASVEGIVNNHVFSMEGFPGKGNVLFGNQLMQIRVTKGGPLPFAFDIVSI  66

Qy   64 QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVK 123
        ||| |:::| :||:|| ||::||| :|| |: | :||:|| : | ||:|| : | || |:
Db   67 AFQYGNRTFTKYPDDIADYFVQSFPAGFFYERNLRFEDGAIVDIRSDISLEDDKFHYKVE 126

Qy  124 FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
        :  ||||| ||||||   :| | || :  ||||| |  | || ||| ||| : |   || :
Db  127 YRGNGFPSNGPVMQKAILGMEPSFEVVVMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS 186

Qy  184 KKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
        | ||  ||| | : || :|  ||| :: | |:|
Db  187 KGGVKEFPEYHFIHHRLEKT-YVEEGSFVBQHE 218
```

```
RESULT 11
US-11-175-766-32
; Sequence 32, Application US/11175766
; Publication No. US20050272111A1
; GENERAL INFORMATION:
;   APPLICANT: Bryan, Bruce
;   APPLICANT: Szczepaniak, William
;   APPLICANT: Szent-Gyorgyi, Christopher
;   TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
;   TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
;   TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
;   FILE REFERENCE: 24729-128
;   CURRENT APPLICATION NUMBER: US/11/175,766
;   CURRENT FILING DATE: 2005-07-06
;   PRIOR APPLICATION NUMBER: US/09/808,898
;   PRIOR FILING DATE: 2001-03-15
;   PRIOR APPLICATION NUMBER: 09/277,716
;   PRIOR FILING DATE: 1999-03-26
;   PRIOR APPLICATION NUMBER: 60/189,691
;   PRIOR FILING DATE: 2000-03-15
;   PRIOR APPLICATION NUMBER: 08/597,274
;   PRIOR FILING DATE: 1996-02-06
;   PRIOR APPLICATION NUMBER: 08/908,909
;   PRIOR FILING DATE: 1997-08-08
;   PRIOR APPLICATION NUMBER: 08/990,103
;   PRIOR FILING DATE: 1997-12-12
;   NUMBER OF SEQ ID NOS: 33
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 32
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Ptilosarcus gurneyi
US-11-175-766-32

  Query Match       39.0%; Score 473.5; DB 7; Length 238;
  Best Local Similarity   45.5%; Pred. No. 8.3e-39;
  Matches   97; Conservative   40; Mismatches   73; Indels   3; Gaps   3;

Qy    5 KNV-IKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP  63
        ||| :|| |::|| : | | ||||:|||| ::||||||||||| | |||||||||:||:|:
Db    7 KNTGLKEIMSAKASVEGIVNNHVFSMEGFPGKGNVLFGNQLMQIRVTKGGPLPFAFDIVSI  66

Qy   64 QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVK 123
        ||| |:::| :||:|| ||::||| :|| |: | :||:|| : | ||:|| : | || |:
Db   67 AFQYGNRTFTKYPDDIADYFVQSFPAGFFYERNLRFEDGAIVDIRSDISLEDDKFHYKVE 126

Qy  124 FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
        :  ||||| ||||||   :| | || :  ||||| |  | || ||| ||| : |   || :
Db  127 YRGNGFPSNGPVMQKAILGMEPSFEVVVMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRS 186

Qy  184 KKPV-QLPGYYYVDSKLDITSHNEDYTIVEQYE 215
        | ||  ||| | : || :|  ||| :: | |:|
Db  187 KGGVKEFPEYHFIHHRLEKT-YVEEGSFVBQHE 218
```

```
RESULT 12
US-11-179-411-27
; Sequence 27, Application US/11179411
; Publication No. US20050266491A1
; GENERAL INFORMATION:
;   APPLICANT: Bryan, Bruce
;   APPLICANT: Szczepaniak, William
;   APPLICANT: Szent-Gyorgyi, Christopher
;   TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
;   TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
;   TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
;   FILE REFERENCE: 24729-128
;   CURRENT APPLICATION NUMBER: US/11/179,411
;   CURRENT FILING DATE: 2005-07-12
;   PRIOR APPLICATION NUMBER: US/09/808,898
;   PRIOR FILING DATE: 2001-03-15
;   PRIOR APPLICATION NUMBER: 60/189,691
;   PRIOR FILING DATE: 2000-03-15
;   PRIOR APPLICATION NUMBER: 09/277,716
;   PRIOR FILING DATE: 1999-03-26
;   PRIOR APPLICATION NUMBER: 08/757,046
;   PRIOR FILING DATE: 1996-11-25
;   PRIOR APPLICATION NUMBER: 08/597,274
;   PRIOR FILING DATE: 1996-02-06
;   PRIOR APPLICATION NUMBER: 08/908,909
;   PRIOR FILING DATE: 1997-08-08
;   PRIOR APPLICATION NUMBER: 08/990,103
;   PRIOR FILING DATE: 1997-12-12
;   NUMBER OF SEQ ID NOS: 33
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 27
;   LENGTH: 233
;   TYPE: PRT
;   ORGANISM: Renilla reniformis
US-11-179-411-27

  Query Match       34.8%; Score 422.5; DB 7; Length 233;
  Best Local Similarity   40.5%; Pred. No. 6.9e-34;
  Matches   90; Conservative   38; Mismatches   93; Indels   1; Gaps   1;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
        |  |||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MDLAKLGLKEVMPTKINLEGLVGDHAFSMEGVGEGNILEGTQEVKISVTKGAPLPFAFDI  60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        :| |||| | :::| :||:|| ::|||||||:|:|| || || : | ||:||:: | |||
Db   61 VSVAFSYGNRAYTGYPBEISDYFLQSFPEGFTYERNIRYQDGGTAIVKSDISLEDDKFIV 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        :|  :  ||||| ||||||    |     : ||||||| | :  | |||| ||| : |
Db  121 NVDFKAKDLRRMGPVMQQDIVGMQPSYESMYTNVTSVIGECIIAFKLQTGKHFTYHMRTV 180
```

Qy 181 YMAKKPVQ-LPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
            |||||| ||| ||||||||||||||||||| |||
Db 181 YKSKKPVETMPLYHFIQHRLVKTNVDTASGVVVQHETAIAAH 222

RESULT 13
US-11-175-766-27
; Sequence 27, Application US/11175766
; Publication No. US20050272111A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/175,766
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 08/757,046
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 08/597,274
; PRIOR FILING DATE: 1996-02-06
; PRIOR APPLICATION NUMBER: 08/908,909
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 08/990,103
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Renilla reniformis
US-11-175-766-27

Query Match   34.8%; Score 422.5; DB 7; Length 233;
Best Local Similarity 40.5%; Pred. No. 6.9e-34;
Matches 90; Conservative 38; Mismatches 93; Indels 1; Gaps 1;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
Db   1 MDLAKLGLKEVWPTKINLEGLVGDHAFSMEGVGEGNILEGTQEVKISVTKGAPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
Db   61 VSVAFSYGNRAVTGYPEEISDYFLQSFPEGFTYERNIRYQDGGTAIVKSDISLEDGKFIV 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Db 121 NVDFKADLRRMGPVMQQDIVGMQPSYESMTNVTSVIGECIIAFKLQTGKHFTYHMRTV 180

Qy 181 YMAKKPVQ-LPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
Db 181 YKSKKPVETMPLYHFIQHRLVKTNVDTASGVVVQHETAIAAH 222

RESULT 14
US-11-179-411-33
; Sequence 33, Application US/11179411
; Publication No. US20050266491A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/179,411
; CURRENT FILING DATE: 2005-07-12
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 08/757,046
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 08/597,274
; PRIOR FILING DATE: 1996-02-06
; PRIOR APPLICATION NUMBER: 08/908,909
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 08/990,103
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Renilla Reniformis mutein
US-11-179-411-33

Query Match   33.8%; Score 410.5; DB 7; Length 233;
Best Local Similarity 40.1%; Pred. No. 9.9e-33;
Matches 89; Conservative 37; Mismatches 95; Indels 1; Gaps 1;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
Db   1 MDLAKLGLKEVWPTKINLEGLVGDHAFSMEGVGEGNILEGTQEVKISVTKGAPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
Db   61 VSVAFSYGNRAVTGYPEEISDYFLQSFPEGFTYERNIRYQDGGTAIVDSDISLEDGKFIV 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Db 121 NVDFKADDLRDMGPVMQQDIVGMQPSYESMTNVTSVIGECIIAFKLQTGKHFTYHMRTV 180

Qy 181 YMAKKPVQ-LPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
Db 181 YKSKKPVETMPLYHFIQHDLVKTNVDTASGVVVQHETAIAAH 222

RESULT 15
US-11-175-766-33
; Sequence 33, Application US/11175766
; Publication No. US20050272111A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/175,766
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 08/757,046
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 08/597,274
; PRIOR FILING DATE: 1996-02-06
; PRIOR APPLICATION NUMBER: 08/908,909
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 08/990,103
; PRIOR FILING DATE: 1997-12-12

```
;  PRIOR FILING DATE: 1997-12-12
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
;
;  SEQ ID NO 33
;  LENGTH: 233
;  TYPE: PRT
;  ORGANISM: Renilla Reniformis mutein
US-11-175-766-33
```

```
Query Match             33.8%;   Score 410.5;   DB 7;   Length 233;
Best Local Similarity   40.1%;   Pred. No. 9.9e-33;
Matches   89;   Conservative   37;   Mismatches   95;   Indels   1;   Gaps   1;


Qy      1 MRSSKNVIKEFMRFKVRMEGGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            : | |   |    |::|  | ||    ||| |   |  | |  ||   |||| || :|
Db      1 MDLAKLGLKEVMPTKINLEGLVGDHAFSMEGGVGEGNILEGTQEVKISVTKGAPLPFAFDI 60

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            :| |||:  |  | : | |||||| : ||||| | |:|| | || :::||    |::|
Db     61 VSVAFSYGNRAYTGYPEEISDYFLQSFPBGFTYERNIRYQDGGTAIVDSDISLEDGKFIV 120

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            |  | : | |:|      |  :  :|  |        |:  | :  |
Db    121 NVDFKADDLRDMGPVMQDIVGMQPSYESMYINVTSVIGECIIAFKLQTGKDFTYHMRTV 180

Qy    181 YMAKKPVQ-LPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
            |  ||||   |  |||  |   : :: :: |:::   :|
Db    181 YKSKKPVETMPLYHFIQHDLVKINVDTASGYVVQHETAIAAH 222
```

THIS PAGE BLANK (USPTO)

OM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2006 Compugen Ltd.
GenCore version 5.1.6

Run on: January 12, 2006, 09:05:37 ; Search time 3691 Seconds
(without alignments)
3465.125 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Xgapop 10.0 ; Xgapext 0.5
Fgapop 6.0 ; Fgapext 0.5
Delop 6.0 ; Delext 7.0

Searched: 5883141 seqs, 2842172565J residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat_10012006_162408_18921/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1.1_7415 @runat_10012006_162408_18921 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=100 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_by:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	678	6	AR183915	AR183915 Sequence
2	1214	100.0	678	6	AX172854	AX172854 Sequence
3	1214	100.0	678	6	AX207715	AX207715 Sequence
4	1214	100.0	678	6	AX233581	AX233581 Sequence
5	1214	100.0	678	6	AX370404	AX370404 Sequence
6	1214	100.0	678	2	AX824725	AX824725 Sequence
7	1214	100.0	859	6	AF168419	AF168419 Discosoma
8	1214	100.0	859	6	AR636082	AR636082 Sequence
9	1214	100.0	859	6	AR670157	AR670157 Sequence
10	1214	100.0	681	6	AX463698	AX463698 Sequence
11	1214	100.0	859	11	AX463702	AX463702 Sequence
12	1210	99.7	713	6	AX823860	AX823860 Sequence
13	1210	99.7	723	6	AR527331	AR527331 Sequence
14	1210	99.7	2721	6	CQ882115	CQ882115 Sequence
15	1210	99.7	2772	6	CQ882117	CQ882117 Sequence
16	1210	99.7	4692	6	AX463702	AX463702 Sequence
17	1210	99.7	6893	6	AX233860	AX233860 Sequence
18	1210	99.7	8811	6	AY569779	AY569779 Cloning v
19	1210	99.7	9320	6	AX663075	AX663075 Sequence
20	1207	99.4	12404	11	AY569780	AY569780 Cloning v
21	1196	98.5	678	6	AX370406	AX370406 Sequence
22	1196	98.5	666	6	AX348043	AX348043 Sequence
23	1196	98.5	711	6	AX353910	AX353910 Sequence
24	1193	98.3	678	2	AY679107	AY679107 Discosoma
25	1191	98.1	678	6	AF545828	AF545828 Discosoma
26	1191	98.1	678	6	AX370408	AX370408 Sequence
27	1191	98.1	921	2	AX824732	AX824732 Sequence
28	1191	98.1	675	6	AY679106	AY679106 Discosoma
29	1186	97.7	1050	6	AX824731	AX824731 Sequence
30	1186	97.7	4488	6	AX666133	AX666133 Sequence
31	1186	97.7	4546	6	CQ849509	CQ849509 Sequence
32	1186	97.7	4555	11	AY818375	AY818375 Cloning v
33	1186	97.7	4570	11	DQ005468	DQ005468 Cloning v
34	1186	97.7	4829	11	AY818373	AY818373 Cloning v
35	1186	97.7	4906	11	AY640628	AY640628 siRNA
36	1186	97.7	4905	11	AY640625	AY640625 siRNA
37	1186	97.7	5311	11	AJ851284	AJ851284 Cloning v
38	1186	97.7	6423	11	AY613997	AY613997 Cloning v
39	1186	97.7	6990	11	AY640634	AY640634 siRNA vec
40	1186	97.7	7058	11	AY640633	AY640633 siRNA vec
41	1186	97.7	7147	11	AY640630	AY640630 siRNA
42	1186	97.7	7616	6	CQ849511	CQ849511 Sequence
43	1186	97.7	9468	6	CS018252	CS018252 Sequence
44	1186	97.7	10140	6	CS018253	CS018253 Sequence
45	1186	97.7	10522	6	CS018254	CS018254 Sequence

ALIGNMENTS

RESULT 1
AR183915

```
LOCUS       AR183915             678 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 6 from patent US 6342379.
ACCESSION   AR183915
VERSION     AR183915.1  GI:20227884
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 678)
AUTHORS     Tsien,R.Y. and Gonzalez,J.E. III.
TITLE       Detection of transmembrane potentials by optical methods
JOURNAL     Patent: US 6342379-A 6 29-JAN-2002;
FEATURES             Location/Qualifiers
source          1..678
                /organism="unknown"
                /mol_type="unassigned DNA"
ORIGIN
```

Alignment Scores:
Pred. No.:   1.4e-114
Score: 1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 6

Length: 678
Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

Left column

US-10-006-922A-12 (1-225) x AR183915 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA  60

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAGGGCCATACGAAGGC  120

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaIlePro  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCATACCA  240

Qy    81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAAATGGGAAAGGTCATGAACTTTGAA  300

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GACGGTGGCGTTGTTACTGTAACCCAGGATTCCAGTTTGCAAGACGGTTGTTTCATCTAC  360

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTTTGAAAGGAGAG  480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ATTCATAAGGCTCTGAAGCTTAAGGATGGTGGTCATTACTTGGTTGAATTCAAAAGTATT  540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TACATGGCAAAGAAGCCTGTCCAGCTACCAGGTTACTACTATGTTGACTCCAAACTGGAT  600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660

Qy   221  HisHisLeuPheLeu  225
          |||||||||||||||
Db   661  CACCATCTGTTCCTT  675
```

RESULT 2

LOCUS AX172854   678 bp   DNA   linear   PAT 03-JUL-2001
DEFINITION Sequence 6 from Patent WO0142211.
ACCESSION AX172854
VERSION AX172854.1   GI:14597903
KEYWORDS .
SOURCE Discosoma sp.
  ORGANISM Discosoma sp.
    Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
    Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE 1
  AUTHORS Tsien,R.Y. and Gonzalez,J.E.
  TITLE Detection of transmembrane potentials by optical methods
  JOURNAL Patent: WO 0142211-A 6 14-JUN-2001;
    The Regents of the University of California (US)
FEATURES   Location/Qualifiers
  source   1..678
    /organism="Discosoma sp."
    /mol_type="unassigned DNA"
    /db_xref="taxon:86600"
ORIGIN

Right column

Alignment Scores:

Pred. No.:	1.4e-114	Length:	678
Score:	1214.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6	Indels:	0
DB:		Gaps:	0

US-10-006-922A-12 (1-225) x AX172854 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA  60

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAGGGCCATACGAAGGC  120

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaIlePro  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCATACCA  240

Qy    81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAAATGGGAAAGGTCATGAACTTTGAA  300

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GACGGTGGCGTTGTTACTGTAACCCAGGATTCCAGTTTGCAAGACGGTTGTTTCATCTAC  360

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTTTGAAAGGAGAG  480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ATTCATAAGGCTCTGAAGCTTAAGGATGGTGGTCATTACTTGGTTGAATTCAAAAGTATT  540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TACATGGCAAAGAAGCCTGTCCAGCTACCAGGTTACTACTATGTTGACTCCAAACTGGAT  600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660

Qy   221  HisHisLeuPheLeu  225
          |||||||||||||||
Db   661  CACCATCTGTTCCTT  675
```

RESULT 3

LOCUS AX207715   678 bp   DNA   linear   PAT 31-AUG-2001
DEFINITION Sequence 13 from Patent WO0157242.
ACCESSION AX207715
VERSION AX207715.1   GI:15422399
KEYWORDS .
SOURCE Discosoma sp.
  ORGANISM Discosoma sp.
    Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
    Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE 1
  AUTHORS Stack,J.H., Whitney,M., Cubitt,A.B. and Pollok,B.A.
  TITLE Methods of protein destabilization and uses thereof

JOURNAL   Patent: WO 0157242-A 13 09-AUG-2001;
             Aurora Biosciences Corporation (US)
FEATURES   Location/Qualifiers
source   1..678
             /organism="Discosoma sp."
             /mol_type="unassigned DNA"
             /db_xref="taxon:86600"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.4e-114   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 100.00%   Indels: 0
                          6   Gaps: 0

US-10-006-922A-12 (1-225) x AX207715 (1-678)

Qy   1   MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA   60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGGAAGGC   120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAATACCGTAAAGCTTAAGGTAACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATT   180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA   240

Qy   81 AspThrTyrLysLeuLysPheProGluGlyGlyProValMetGlnAsnPheGlu   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTATAAAAAAGCTGTCATTTCCTGAAGGAGGTCCAGTTATGCAGAACTTTGAA   300

Qy   101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC   360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA   420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAGGGTGAA   480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT   540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT   600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluIleTyrGluArgThrGluGlyArg   220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGATATATGAAAGAACCGAGGGACGC   660

Qy   221 HisHisLeuPheLeu   225
          ||||||||||||||
Db   661 CACCATCGTTCCTT   675

KEYWORDS   .
SOURCE   Discosoma sp.
  ORGANISM   Discosoma sp.
              Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
              Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Nelson,D., Zamaira,E. and Tsien,R.
  TITLE   Modified fluorescent proteins
  JOURNAL   Patent: WO 0162919-A 5 30-AUG-2001;
              Aurora Biosciences Corporation (US)
FEATURES   Location/Qualifiers
source   1..678
              /organism="Discosoma sp."
              /mol_type="unassigned DNA"
              /db_xref="taxon:86600"
              /note="'red'"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.4e-114   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 100.00%   Indels: 0
                          6   Gaps: 0

US-10-006-922A-12 (1-225) x AX233581 (1-678)

Qy   1   MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA   60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGGAAGGC   120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAATACCGTAAAGCTTAAGGTAACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATT   180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA   240

Qy   81 AspThrTyrLysLeuLysPheProGluGlyGlyProValMetGlnAsnPheGlu   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTATAAAAAAGCTGTCATTTCCTGAAGGAGGTCCAGTTATGCAGAACTTTGAA   300

Qy   101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC   360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA   420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAGGGTGAA   480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT   540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT   600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluIleTyrGluArgThrGluGlyArg   220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGATATATGAAAGAACCGAGGGACGC   660

Qy   221 HisHisLeuPheLeu   225

RESULT 4
LOCUS   AX233581   678 bp   DNA   linear   PAT 11-SEP-2001
DEFINITION   Sequence 5 from Patent WO0162919.
ACCESSION   AX233581
VERSION   AX233581.1   GI:15593305

RESULT 5

LOCUS   AX370404   678 bp   DNA   linear   PAT 16-FEB-2002
DEFINITION   Sequence 1 from Patent WO0196373.
ACCESSION   AX370404
VERSION   AX370404.1 GI:18857490
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM   synthetic construct
               other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Fradkov,A.F. and Terskikh,A.
  TITLE   Fluorescent timer proteins and methods for their use
  JOURNAL   Patent: WO 019373-A 1 20-DEC-2001;
               Clontech Laboratories Inc. (US)
FEATURES   Location/Qualifiers
     source   1..678
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="variant of sequence from Discosoma sp."
ORIGIN

US-10-006-922A-12 (1-225) x AX370404 (1-678)

Alignment Scores:
  Pred. No.:
  Score: 1214.00   Length: 678
  Percent Similarity: 100.00%   Matches: 225
  Best Local Similarity: 100.00%   Conservative: 0
  Query Match: 6   Mismatches: 0
                                        Indels: 0
  DB: Gaps: 0

Qy   1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCGTCTCCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspThrTyrLysLeuSerPheProGluGlyGlyProValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACACACAGAAGCTGTCCTTCCCCGAGGGCGGCCCCGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTCCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluIleTyrGluArgGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCTACGAGCGCGGCCGC 660

Qy   221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   661 CACCACCTGTTCCTT 675

RESULT 6

LOCUS   AX824725   678 bp   DNA   linear   PAT 11-DEC-2003
DEFINITION   Sequence 7 from Patent WO02068459.
ACCESSION   AX824725
VERSION   AX824725.1 GI:39750591
KEYWORDS   .
SOURCE   Discosoma sp.
  ORGANISM   Discosoma sp.
               Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
               Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Non aggregating fluorescent proteins and methods for using the same
  TITLE
  JOURNAL   Patent: WO 02068459-A 7 06-SEP-2002;
FEATURES   Location/Qualifiers
     source   1..678
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"
ORIGIN

US-10-006-922A-12 (1-225) x AX824725 (1-678)

Alignment Scores:
  Pred. No.:
  Score: 1214.00   Length: 678
  Percent Similarity: 100.00%   Matches: 225
  Best Local Similarity: 100.00%   Conservative: 0
  Query Match: 6   Mismatches: 0
                                        Indels: 0
  DB: Gaps: 0

Qy   1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCGTCTCCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspThrTyrLysLeuSerPheProGluGlyGlyProValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACACACAGAAGCTGTCCTTCCCCGAGGGCGGCCCCGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTCCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

RESULT 7

```
Db  AF168419
Qy  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCGACGACGGCGTGCTGAAGGGCGAG 480
Qy  161 IleHisHisLysAlaLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy  481 ATCCACCAAGGCCCTGAAGCTGGAAGGACGGCGGGGGCACTACCTGGTGGAGTTCAAGTCCATC 540
Qy  181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy  541 TACATGGCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660
Qy  221 HisHisLeuPheLeu 225
Qy  661 CACCACCTGTTCCTG 675
```

```
LOCUS       AF168419      859 bp    mRNA    linear  INV 27-JUL-2001
DEFINITION  Discosoma sp. fluorescent protein FP583 mRNA, complete cds.
ACCESSION   AF168419
VERSION     AF168419.2  GI:7105733
KEYWORDS    .
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
            Markelov,M.L. and Lukyanov,S.A.
  TITLE     Fluorescent proteins from nonbioluminescent Anthozoa species
  JOURNAL   Nat. Biotechnol. 17 (10), 969-973 (1999)
  PUBMED    10504696
REFERENCE   2  (bases 1 to 859)
  AUTHORS   Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
            Markelov,M.L. and Lukyanov,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1999) Institute of Bioorganic Chemistry RAS,
            Miklukho-Maklaya 16/10, Moscow 117871, Russia
REFERENCE   3  (bases 1 to 859)
  AUTHORS   Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
            Markelov,M.L. and Lukyanov,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Institute of Bioorganic Chemistry RAS,
            Miklukho-Maklaya 16/10, Moscow 117871, Russia
REMARK      Sequence update by submitter
COMMENT     On Feb 25, 2000 this sequence version replaced gi:6090866.
            Sequence update by submitter replaced gi:6090866.
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="Discosoma sp."
                     /mol_type="mRNA"
                     /db_xref="taxon:86600"
     CDS             54..731
                     /codon_start=1
                     /note="GFP-like protein; orange-red"
                     /product="fluorescent protein FP583"
                     /protein_id="AAF03369.1"
                     /db_xref="GI:6090867"
                     /translation="MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
ORIGIN
```

Alignment Scores:

```
Pred. No.:           1.
Score:               1214.00
Percent Similarity:  100.00%
Best Local Similarity: 100.00%
Query Match:         100.00%
```

```
                     1.87e-114       Length:    859
                                     Matches:   225
                                     Conservative: 0
                                     Mismatches: 0
                                     Indels:    0
```

RESULT 8

```
DB:   US-10-006-922A-12 (1-225) x AF168419 (1-859)    2    Gaps:    0
```

```
Qy  1   MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTAATGAACTTTGAA 353
Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCAGTTTGCAGATGGCGTTTCATCTAC 413
Qy  41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTCCTTCCGGATGGACCCGTTATGCAAAAGAAGACA 473
Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTGTATCCTCGTGATGGCTGTTGAAAAGGAGAG 533
Qy  81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCACTACCTAGTTGAATTCAAAAGTATT 593
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT 653
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  714 CACCACCTGTTCCTT 728
Qy  161 IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy  221 HisHisLeuPheLeu 225
```

```
LOCUS       AR636082      859 bp    DNA    linear  PAT 14-FEB-2005
DEFINITION  Sequence 11 from patent US 6852849.
ACCESSION   AR636082
VERSION     AR636082.1  GI:59795931
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Tsien,R.Y. and Campbell,R.E.
  TITLE     Non-oligomerizing tandem fluorescent proteins
  JOURNAL   Patent: US 6852849-A 11 08-FEB-2005;
            The Regents of the University of California; Oakland, CA
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                    1.87e-114     Length:        859
                                        Matches:       225
Percent Similarity:       100.00%       Conservative:  0
                                        Mismatches:    0
Best Local Similarity:    100.00%       Indels:        0
Query Match:              6             Gaps:          0
DB:

US-10-006-922A-12 (1-225) x AR636082 (1-859)

Qy    1 MetArgSerSerLysAsnValIleLysGluLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTCGCATGAGAAGGA 113

Qy   21 ThrValAsnGlyHisGlyPheGluIleLeuGluGluGlyGluGlyAspGlyProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTTCAATGGGCACCAGAGTTTGAAATAGAAGGCGAAGGCGAAGGAGGAGGACCCATACGAAGGC 173

Qy   41 HisAsnThrValLysGlyLeuLysValAlaThrLysGlyLeuProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGGTTAAGGTAGCCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACGTGTACCCAGGATTCCAGTTTGCAGGATGGCGTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTCTGAAAGGAGAG 533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy  221 HisHisLeuPheLeu 225
         |||||||||||||||
Db  714 CACCATCTGTTCCTT 728
```

RESULT 9

```
LOCUS       AR670157        859 bp    DNA    linear    PAT 13-JUN-2005
DEFINITION  Sequence 11 from patent US 6900304.
ACCESSION   AR670157
VERSION     AR670157.1  GI:67608896
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Tsien,R.Y., Ting,A.Y. and Zhang,J.
  TITLE     Emission ratiometric indicators of phosphorylation
  JOURNAL   Patent: US 6900304-A 11 31-MAY-2005;
            The Regents of the University of California; Oakland, CA
```

```
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                    1.87e-114     Length:        859
                                        Matches:       225
Percent Similarity:       100.00%       Conservative:  0
                                        Mismatches:    0
Best Local Similarity:    100.00%       Indels:        0
Query Match:              6             Gaps:          0
DB:

US-10-006-922A-12 (1-225) x AR670157 (1-859)

Qy    1 MetArgSerSerLysAsnValIleLysGluLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTCGCATGAGAAGGA 113

Qy   21 ThrValAsnGlyHisGlyPheGluIleLeuGluGluGlyGluGlyAspGlyProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTTCAATGGGCACCAGAGTTTGAAATAGAAGGCGAAGGCGAAGGAGGAGGACCCATACGAAGGC 173

Qy   41 HisAsnThrValLysGlyLeuLysValAlaThrLysGlyLeuProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGGTTAAGGTAGCCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACGTGTACCCAGGATTCCAGTTTGCAGGATGGCGTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTCTGAAAGGAGAG 533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy  221 HisHisLeuPheLeu 225
         |||||||||||||||
Db  714 CACCATCTGTTCCTT 728
```

RESULT 10

```
LOCUS       AX463698        859 bp    DNA    linear    PAT 15-JUL-2002
DEFINITION  Sequence 12 from Patent WO0248338.
ACCESSION   AX463698
VERSION     AX463698.1  GI:21886457
KEYWORDS
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
```

```
REFERENCE   1
  AUTHORS   Lichtenberg-Prat.H.
  TITLE     Yeast strain for testing the geno- and cytotoxicity of complex
            environmental contamination
  JOURNAL   Patent: WO 0248336-A 12 20-JUN-2002;
            Lichtenberg-Prat, Hella (DE)

FEATURES             Location/Qualifiers
  source     1..859
             /organism="Discosoma sp."
             /mol_type="unassigned DNA"
             /db_xref="taxon:86600"
  CDS        54..731
             /codon_start=1
             /note="unnamed protein product"
             /protein_id="CAD42147.1"
             /db_xref="GI:21886458"
             /translation="MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
             KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
             GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
             IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
             GRHHLFL"
ORIGIN

Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
```

US-10-006-922A-12 (1-225) x AX463698 (1-859)

```
Alignment Scores:
Pred. No.:
Score:                    1.87e-114        Length:          859
                          1214.00          Matches:         225
Percent Similarity:       100.00%          Conservative:    0
Best Local Similarity:    100.00%          Mismatches:      0
Query Match:              6                Indels:          0
DB:                                        Gaps:            0
```

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGGATGGAAGGA 113
Qy   21 ThrValAsnGlyHisGluPheIleGluGlyGluGlyGluGlyArgProTyrGluGly     40
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGCGAAGGCCATACGAAGGC     173
Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353
Qy  101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr    120
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTCATCTAC  413
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAGAAGACA  473
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTGAAAAGGGAGAG 533
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT 653
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713
Qy  221 HisHisLeuPheLeu 225
Db  714 CACCATCTGTTCCTT 728
```

RESULT 11
```
LOCUS       AX233584                681 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 8 from Patent WO0162919.
ACCESSION   AX233584
VERSION     AX233584.1  GI:15593307
KEYWORDS
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Nelson,D., Zamaira,E. and Tsien,R.
  TITLE     Modified fluorescent proteins
  JOURNAL   Patent: WO 0162919-A 8 30-AUG-2001;
            Aurora Biosciences Corporation (US)

FEATURES             Location/Qualifiers
  source     1..681
             /organism="synthetic construct"
             /mol_type="unassigned DNA"
             /db_xref="taxon:32630"
             /note="Mutant Anthozoan red fluorescent protein"
  CDS        1..681
             /codon_start=1
             /note="unnamed protein product"
             /transl_table=11
             /protein_id="CAC69734.1"
             /db_xref="GI:15593308"
             /translation="MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNT
             VKLKVTKGGPLPFAWDILSPQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFED
             GGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKG
             EIHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERT
             EGRHHLFL"
ORIGIN
```

US-10-006-922A-12 (1-225) x AX233584 (1-681)

```
Alignment Scores:
Pred. No.:
Score:                    3.61e-114        Length:          681
                          1210.00          Matches:         224
Percent Similarity:       100.00%          Conservative:    1
Best Local Similarity:    99.56%           Mismatches:      0
Query Match:              6                Indels:          0
DB:                       99.67%           Gaps:            0
```

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
Db    4 GTGAGGAGGCAGCAAGAACCTGATCAAGGAGTTCATGAGGTTCAAGGTGCGCATGGAGGGC 63
Qy   21 ThrValAsnGlyHisGluPheIleGluGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db   64 ACCGTGAACGGGCCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123
Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  124 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183
Qy   61 LeuSerProGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro       80
Db  184 CTGAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  244 GACTACAAGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 303
```

RESULT 12

LOCUS	AX233627　713 bp　DNA　linear　PAT 11-SEP-2001
DEFINITION	Sequence 51 from Patent WO0162919.
ACCESSION	AX233627
VERSION	AX233627.1　GI:15593330
KEYWORDS	.
SOURCE	synthetic construct
ORGANISM	synthetic construct
	other sequences; artificial sequences.
REFERENCE	1
AUTHORS	Nelson,D., Zamaira,E. and Tsien,R.
TITLE	Modified fluorescent proteins
JOURNAL	Patent: WO 0162919-A 51 30-AUG-2001;
	Aurora Biosciences Corporation (US)
FEATURES	Location/Qualifiers
source	1..713
	/organism="synthetic construct"
	/mol_type="unassigned DNA"
	/db_xref="taxon:32630"
	/note="Mutant Anthozoan red fluorescent protein"
ORIGIN	

Alignment Scores:

Pred. No.:	1		
Score:	1210.00		
Percent Similarity:	100.00%	Matches:	224
Best Local Similarity:	99.56%	Conservative:	1
Query Match:	99.67%	Mismatches:	0
DB:	6	Indels:	0
		Gaps:	0

US-10-006-922A-12 (1-225) x AX233627 (1-713)

```
Qy    1 MetArgSerLysAsnValIleLeuLysGlyGluPheMetArgPheLysValArgMetGluGly 20
        :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MetArgSerLysAsnValIleLeuLysGlyGluPheMetArgPheLysValArgMetGluGly 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 141

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 201
```

RESULT 13

LOCUS	AR527331　723 bp　DNA　linear　PAT 08-OCT-2004
DEFINITION	Sequence 1 from patent US 6723537.
ACCESSION	AR527331
VERSION	AR527331.1　GI:53914309
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
	Unclassified.
REFERENCE	1　(bases 1 to 723)
AUTHORS	Peelle,B.
TITLE	Directed evolution of protein in mammalian cells
JOURNAL	Patent: US 6723537-A 1 20-APR-2004;
	Rigel Pharmaceuticals, Incorporated and Becton, Dickinson and
	Company; South San Francisco, CA
FEATURES	Location/Qualifiers
source	1..723
	/organism="unknown"
	/mol_type="genomic DNA"
ORIGIN	

Alignment Scores:

Pred. No.:	1		
Score:	1210.00		
Percent Similarity:	100.00%	Matches:	224
Best Local Similarity:	99.56%	Conservative:	1
Query Match:	99.67%	Mismatches:	0
DB:	6	Indels:	0
		Gaps:	0

US-10-006-922a-12 (1-225) x AR527331 (1-723)

```
Qy    1 MetArgSerLysAsnValIleLeuLysGlyGluPheMetArgPheLysValArgMetGluGly 20
        :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4 GTGCGGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63
```

ORIGIN
Alignment Scores:

FEATURES
          source

          REFERENCE
          AUTHORS

          TITLE

          JOURNAL

          ORGANISM
          KEYWORDS
          VERSION
          ACCESSION
          DEFINITION
          LOCUS

RESULT 14

```
Location/Qualifiers
1..2721
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description de la s quence artificielle : s quence
de fusion"
```

1
Chavancy,G., Couble,P., Durand,B., Grenier,A.M., Horard,B.,
Julien,E., Mauchamp,B., Nony,P., Prudhomme,J.C. and Royer,C.
Centre National De La Recherche Scientifique-CNRS (FR); UNIVERSITE
CLAUDE BERNARD - LYON 1 (FR); Institut National de la Recherche
Agronomique (INRA) (FR)
Patent: WO 200408345-A1 30-SEP-2004;
the posterior silk glands of a lepidoptera and application thereof
Nucleic acid controlling the expression of a useful polypeptide in
other sequences; artificial sequences.
synthetic construct
synthetic construct
CQ882115.1   GI:54034825
CQ882115
CQ882115   Sequence 1 from Patent WO2004083445.
CQ882115   2721 bp   DNA   linear   PAT 11-OCT-2004

Db   664 CACCACCTGTTCCTG 678
Qy   221 HisHisHisLeuPheLeu 225

Db   604 ATCACCTCCCACAACGAGACTACACCATCGTGGAGAGCAGTACGACGAGGCCACGAGGGCCGC 663
Qy   201 IleThrSerHisAsnGluAspThrTyrThrIleValGluGlnTyrProArgAspSerSerIle 220

Db   544 TACATGGCCAAGAAGCCCTGAAGCTGGAGGTTCATCGGGCGTACTACTACGTGGAGTCCAAGCTGGAC 603
Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200

Db   484 ATCGACACAAGGGCCCTGAAGCTGAAGGACGGGCGCCGGCCACTACCTGGTGCTTCAATCTAC 543
Qy   161 IleHisHisAlaLeuLysLeuLysAspGlyHisHisTyrLeuValLeuPheSerSerIle 180

Db   424 ATGGGCTGGGAGGGCCCTGACACCCGGACGGGGCCGGTGGCTGAGTTCAAGGAGGTGGAG 483
Qy   141 MetGlyTrpGluGlyAlaSerThrArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlu 160

Db   364 AAGGTGAAGTTCATCGGGCGTGAAGACTTCCCCGGTAATGCAGGAGCACCCGGCGACATCCC 423
Qy   121 LysValLysPheIleGlyValLysAsnPheProGlyAsnAspGlyProValMetGlnGlyAsnThr 140

Db   304 GACGGGCGGCCGGTGGTGACCGTGACGGAGGGCCGGTGCTTCATCTAC 363
Qy   101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Db   244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCCTTCAAGTGGGAGTGGGAGCGCGGGC 303
Qy   81 AspTyrLysLysLeuSerPheProGluGlyLeuGlyPheLysTrpGluGlyAlaGlyGluArg 100

Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAGAGGTGACCAAGAGGCGGGCCCCCTTCGCCTCTGGGACATC 243
Qy   61 LeuSerProGlnPheGlnTyrGlySerArgGlyValThrLysArgValValLysHisProAlaAspIlePro 80

Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGAGGGGGGGCCCCCTTGCCCTTTGCCTCTGGGACATC 183
Qy   41 HisAsnThrValLysLeuLysValThrLysArgGlyGlyProLeuProPheAlaTrpAspIle 60

Db   64 ACCGTGAACGGCCCACCAGTTCGAGATCGAGGGCGACGGGCGAGGGCGGGGCCCCTTACGAGGGC 123
Qy   21 ThrValAsnGlyHisHisGlyPheGlulleGlulleGluGlyAspGlyAspGlyArgProTyrGluGly 40
```

Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

| | | |
|---|---|---|
| Length: | 2721 | |
| Matches: | 224 | |
| Mismatches: | 1 | |
| Conservative: | 0 | |
| Indels: | 0 | |
| Gaps: | 0 | |

```
1.95e-113
1210.00
100.00%
99.56%
99.67%
6
```

US-10-006-922A-12 (1-225) x CQ882115 (1-2721)

```
Db    2044 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATCGGCCTTCAAGGTGCGCATGGAGGGC 2103
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

Db    2104 ACCGTGAACGGCCACCAGTTCGAGATCGAGGGCGACGGGCGAGGGCGGGGCCCCTTACGAGGGC 2163
Qy    21 ThrValAsnGlyHisHisGlyPheGluIleGluIleGluGlyAspGlyAspGlyArgProTyrGluGly 40

Db    2164 CACAACACCGTGAAGCTGAAGGTGACCAAGAGGGGGGGCCCCCTTGCCCTTTGCCTCTGGGACATC 2223
Qy    41 HisAsnThrValLysLeuLysValThrLysArgGlyGlyProLeuProPheAlaTrpAspIle 60

Db    2224 CTGTCCCCCCAGTTCCAGTACGGCTCCAGAGGTGACCAAGAGGCGGGCCCCCTTCGCCTCTGGGACATC 2283
Qy    61 LeuSerProGlnPheGlnTyrGlySerArgGlyValThrLysArgValValLysHisProAlaAspIlePro 80

Db    2284 GACTACAAGAAGCTGTCCTTCCCCGAGGGCCTTCAAGTGGGAGTGGGAGCGCGGGC 2343
Qy    81 AspTyrLysLysLeuSerPheProGluGlyLeuGlyPheLysTrpGluGlyAlaGlyGluArg 100

Db    2344 GACGGGCGGCCGGTGGTGACCGTGACGGAGGGCCGGTGCTTCATCTAC 2403
Qy    101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Db    2404 AAGGTGAAGTTCATCGGGCGTGAAGACTTCCCCGGTAATGCAGGAGCACCCGGCGACATCCC 2463
Qy    121 LysValLysPheIleGlyValLysAsnPheProGlyAsnAspGlyProValMetGlnGlyAsnThr 140

Db    2464 ATGGGCTGGGAGGGCCCTGACACCCGGACGGGGCCGGTGGCTGAGTTCAAGGAGGTGGAG 2523
Qy    141 MetGlyTrpGluGlyAlaSerThrArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlu 160

Db    2524 ATCGACACAAGGGCCCTGAAGCTGAAGGACGGGCGCCGGCCACTACCTGGTGCTTCAATCTAC 2583
Qy    161 IleHisHisAlaLeuLysLeuLysAspGlyHisHisTyrLeuValLeuPheSerSerIle 180

Db    2584 TACATGGCCAAGAAGCCCTGAAGCTGCCCGGGTACTACTACTACGTGGACTCCAAGCTGGAC 2643
Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200

Db    2644 ATCACCTCCCACAACGAGACTACACCATCGTGGAGAGCAGTACGACGAGGCCACGAGGGCCGC 2703
Qy    201 IleThrSerHisAsnGluAspThrTyrThrIleValGluGlnTyrProArgAspSerSerIle 220

Db    2704 CACCACCTGTTCCTG 2718
Qy    221 HisHisHisLeuPheLeu 225
```

RESULT 15
          LOCUS   CQ882117   2772 bp   DNA   linear   PAT 11-OCT-2004
          DEFINITION   CQ882117   Sequence 3 from Patent WO2004083445.
          ACCESSION   CQ882117
          VERSION   CQ882117.1   GI:54034827
          KEYWORDS
          SOURCE   synthetic construct
          ORGANISM   synthetic construct
                         other sequences; artificial sequences.
          REFERENCE   1
          AUTHORS   Chavancy,G., Couble,P., Durand,B., Grenier,A.M., Horard,B.,
                         Julien,E., Mauchamp,B., Nony,P., Prudhomme,J.C. and Royer,C.
          TITLE   Nucleic acid controlling the expression of a useful polypeptide in
                         the posterior silk glands of a lepidoptera and application thereof
          JOURNAL   Patent: WO 200408345-A3 30-SEP-2004;

FEATURES
source

Centre National De La Recherche Scientifique-CNRS (FR); UNIVERSITE
CLAUDE BERNARD - LYON 1 (FR); Institut National de la Recherche
Agronomique (INRA) (FR)
Location/Qualifiers
1. .2772
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description de la s quence artificielle : s quence
de fusion"

ORIGIN

US-10-006-922A-12 (1-225) x CQ882117 (1-2772)

Alignment Scores:
Pred. No.:
Score: 2e-113   Length: 2772
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
DB: 99.67%   Indels: 0
                        6   Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
Db 2095 GTGCGCCTCTCCAAGAACGTCATCAAGGAGTTCATGCGTTCAAGGTGCGCATGGAGGGC 2154

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db 2155 ACCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2214

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db 2215 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2274

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db 2275 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2334

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 2335 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2394

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 2395 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 2454

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 2455 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTNAATGCAGAAGAAGACC 2514

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 2515 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2574

Qy 161 IleHisLysAlaLeuLysLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Db 2575 ATCCACAAGGCCCTGAAGAAGGACGGCCACTACCTGGTGGAGTTCAAGTCCATC 2634

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 2635 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2694

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 2695 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2754

Qy 221 HisHisLeuPheLeu 225
Db 2755 CACCACCTGTTCCTG 2769

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 12, 2006, 05:49:00 ; Search time 479 Seconds
                                             (without alignments)
                                             3130.597 Million cell updates/sec

               GenCore version 5.1.6
               Copyright (c) 1993 - 2006 Compugen Ltd.

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..............EDYIIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0 , Fgapext 7.0
               Delop   6.0 , Delext   7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

               Listing first 45 summaries

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool_p/US10006922/runat_10012006_162408_18913/app_query.fasta_1.391
 -DB=N_Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LCOPCL=0 -LCOPEXT=0
 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10006922 @CGN_1_1.1096 @runat_10012006_162408_18913 -NCPU=6 -ICPU=3
 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_21:*
               1:   geneseqn1980s:*
               2:   geneseqn1990s:*
               3:   geneseqn2000s:*
               4:   geneseqn2001as:*
               5:   geneseqn2001bs:*
               6:   geneseqn2002as:*
               7:   geneseqn2002bs:*
               8:   geneseqn2003as:*
               9:   geneseqn2003bs:*
              10:   geneseqn2003cs:*
              11:   geneseqn2003ds:*
              12:   geneseqn2004as:*
              13:   geneseqn2004bs:*
              14:   geneseqn2005s:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	678	4	AAH47654	Discosoma
2	1214	100.0	678	4	AAD13053	Discosoma
3	1214	100.0	678	4	AAD11142	Discosoma
4	1214	100.0	678	6	ABA95905	Yeast opt

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

5	1214	100.0	678	6	ABA95922	Aba95922 Yeast opt
6	1214	100.0	678	6	ABA95921	Aba95921 Yeast opt
7	1214	100.0	678	6	ABA95920	Aba95920 Yeast opt
8	1214	100.0	678	6	AAD46278	Aad46278 Discosoma
9	1214	100.0	678	10	AAD28207	Aad28207 Discosoma
10	1214	100.0	678	10	ADC24127	Adc24127 Discosoma
11	1214	100.0	678	10	ADF70404	Adf70404 Discosoma
12	1214	100.0	678	11	ADL46204	Adl46204 Discosoma
13	1214	100.0	678	11	ADN33978	Adn33978 Wild-type
14	1214	100.0	678	12	ADM97768	Adm97768 D sp red
15	1214	100.0	678	12	ADI36420	Adi36420 D sp red
16	1214	100.0	695	3	AAA48743	Aaa48743 Humanised
17	1214	100.0	859	7	AAL47952	Aal47952 Discosoma
18	1214	100.0	859	7	ADY51715	Ady51715 Discosoma
19	1214	100.0	859	8	AAD53432	Aad53432 Discosoma
20	1214	100.0	859	10	AAD61969	Aad61969 Discosoma
21	1214	100.0	859	14	ADX26533	Adx26533 Discosoma
22	1214	100.0	3311	9	ACA62995	Aca62995 Plasmid D
23	1210	99.7	681	10	ADC24134	Adc24134 Discosoma
24	1210	99.7	681	11	ADL46205	Adl46205 Human cod
25	1210	99.7	681	11	ADL46225	Adl46225 Human cod
26	1210	99.7	723	8	ABZ22476	Abz22476 Mammalian
27	1210	99.7	1638	10	ADL18131	Adl18131 RFP;PS(NI
28	1210	99.7	1647	10	ADL18155	Adl18155 RFP:PS(HI
29	1210	99.7	2721	13	ADS75466	Ads75466 Fibrohexa
30	1210	99.7	2721	13	ADS75468	Ads75468 Fibrohexa
31	1210	99.7	4692	6	AAL47954	Aal47954 Modified
32	1210	99.7	4692	10	ACC44640	Acc44640 Vector pD
33	1210	99.7	5436	4	ADI10003	Adi10003 Plasmid_p
34	1210	99.7	6893	10	ADE24111	Ade24111 Proviral
35	1210	99.7	7910	4	AAD09979	Aad09979 pBIT(DHSP
36	1210	99.7	9320	6	ABS56664	Abs56664 Plasmid p
37	1210	99.7	9658	12	ADI04104	Adi04104 Vector pT
38	1210	99.7	9658	12	ADI04103	Adi04103 Vector pT
39	1210	99.7	9678	12	ADI04102	Adi04102 Vector pT
40	1210	99.7	10263	12	ABA95906	Aba95906 Yeast opt
41	1209	99.6	681	6	ABA95907	Aba95907 Yeast opt
42	1207	99.4	678	6	AAD28208	Aad28208 Discosoma
43	1205	99.3	681	4	AAH47656	Aah47656 Anthozoan
44	1203	99.1	7508	13	ADR30814	Adr30814 Zebrafish
45	1197	98.6	6985	10	ACA55359	Aca55359 Transform

ALIGNMENTS

RESULT 1
AAH47654

ID   AAH47654 standard; cDNA; 678 BP.
XX
AC   AAH47654;
XX
DT   30-NOV-2001 (first entry)
XX
DE   Discosoma sp. red drFP583 protein coding sequence.
KW   Fluorescent protein; Anthozoan; fluorescence; marker; FRET; drFP583; ss.
OS   Discosoma sp.
XX
PN   WO200162919-A1.
XX
PA   (AURO-) AURORA BIOSCIENCES CORP.
PI   Nelson D, Zamaira E, Tsien R;
XX
PD   30-AUG-2001.
XX
PF   13-FEB-2001; 2001WO-US004625.
XX
PR   23-FEB-2000; 2000US-0184732P.
XX
PT
DR   WPI; 2001-557704/62.

```
PT  Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
PT  functional red fluorescent proteins, and the encoding nucleic acids, with
PT  key mutations for improving the proteins function.
XX
PS  Disclosure; Page 83; 90pp; English.
XX
CC  The invention provides a nucleic acid encoding functional red fluorescent
CC  protein (II) that differs from the sequence of an Anthozoan red
CC  fluorescent protein by at least one amino acid substitution, and with
CC  different fluorescent properties. The red fluorescent protein of the
CC  invention can be expressed by standard recombinant methodology. (II) are
CC  used a fluorescent markers and FRET partners. It is used for identifying
CC  protein-protein interactions. (II) is also suitable for multiplexed
CC  fluorescent analysis and FRET-based applications using existing Aequorea
CC  fluorescent proteins. (II) has improved brightness, reduced spectral
CC  cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC  The key mutations in the encoding nucleic acids provide improved folding,
CC  brightness, and create (II) with sharper, more defined excitation and
CC  emission peaks when expressed in mammalian cells. The present sequence
CC  represents the coding sequence of a Discosoma sp. drFP583 protein, an
CC  anthozoan fluorescent protein
XX
SQ  Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:
Score: 2.3e-140
Percent Similarity: 100.00%   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 100.00%   Conservative: 0
DB: 4   Mismatches: 0
                                              Indels: 0
                                              Gaps: 0

US-10-006-922A-12 (1-225) x AAH47654 (1-678)

```
Qy   1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTAAGAATGGAAGGA  60

Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACAGAAGGC    120

Qy  41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCGGACATACCA 240

Qy  81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAACGGGTCATGAACTTTGAA 300

Qy 101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361  AAGGTCAAGTTCATTGGCGTTAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy 161  IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluLeuLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221  HisHisLeuPheLeu 225
        |||||||||||||||
Db 661  CACCACCTGTTCCTT 675
```

RESULT 2
AAD13053
```
ID   AAD13053 standard; cDNA; 678 BP.
XX
AC   AAD13053;
XX
DT   16-OCT-2001 (first entry)
XX
DE   Discosoma sp. "red" anthozoa fluorescent protein, drFP583 cDNA.
XX
KW   Protein destabilisation; linker moiety; reporter moiety; disease model;
KW   linear multimerised destabilization domain; -NH-ubiquitin protein endoprotease;
KW   transgenic animal; transgenic plant; disease resistance;
KW   anthozoa fluorescent protein; natural fluorescent protein; ss.
XX
OS   Discosoma sp.
XX
PN   WO200157242-A2.
XX
PI   Stack JH, Whitney M, Cubitt AB, Pollok BA;
XX
DR   WPI; 2001-488890/53.
XX
PA   (AURO-) AURORA BIOSCIENCES CORP.
XX
PR   04-FEB-2000; 2000US-00498098.
XX
PT   02-FEB-2001; 2001WO-US003791.
PT   09-AUG-2001.
XX
PS   Disclosure; Page 110; 171pp; English.
XX
CC   The present invention relates to a method for destabilising a target
CC   protein in a cell. The method comprises a linker moiety which operatively
CC   couples a target protein (a reporter moiety) to a linear multimerised
CC   destabilising domain, which is non-cleavable by a -NH-ubiquitin protein
CC   endoproteases. The method is useful for detecting an activity such as
CC   protease, protein kinase or phosphoprotein phosphatase activity and is
CC   also useful for identifying modulators of these activities. The method is
CC   also useful for developing novel assays for a wide range of post-
CC   translational activities, such as proteolysis, phosphorylation,
CC   dephosphorylation, glycosylation, methylation, sulfation, prenylation,
CC   disulfide bond formation and ADP-ribosylation within cells. The
CC   recombinant DNA molecule of the invention is useful for creating
CC   transgenic animals useful as disease models and transgenic plants with
CC   improved disease resistance or other favourable traits. The present
CC   sequence is Discosoma sp. "red" anthozoa fluorescent protein, drFP483
CC   cDNA which is a natural fluorescent protein used as a reporter moiety in
CC   the exemplification of the invention
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:
Score: 2.3e-140
Percent Similarity: 100.00%   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 100.00%   Conservative: 0
DB: 4   Mismatches: 0
                                              Indels: 0
                                              Gaps: 0

US-10-006-922A-12 (1-225) x AAD13053 (1-678)

```
Qy      1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAGGA 60

Qy     21 ThrValAlaAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyLysProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGGTCAATGGGCACGAGTTTATTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACAATACCGTAAAGCTTAAGGTTAACCAAGGGGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCACTATCTAGTAGAGTTCAAAAGTATT 540

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluIleGluGlnAvaThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTAGAAAAGAACCGAGGGGAGGC 660

Qy    221 HisHisHisLeuPheLeu 225
          ||||||||||||||||||
Db    661 CACCATCTGTTCCTTT 675
```

RESULT 3

PN   WO200142211-A2.
XX
PD   14-JUN-2001.
XX
PF   12-DEC-2000; 2000WO-US033739.
XX

Determining electrical potential across a membrane in biological systems, comprises introducing two reagents, exposing the membrane to light and measuring the energy transfer.

Disclosure; Page 150; 154pp; English.

The patent discloses optical methods and compositions for determining transmembrane potential across biological membranes in living cells. The method of determining the electrical potential across a membrane in a biological system comprises introduction of two reagents, a first reagent comprising a hydrophobic fluorescent ion capable of redistributing from a first face of the membrane to a second face of the membrane in response to membrane potential change and a second reagent that label the first face or the second face of the membrane, which comprises a chromophore capable of undergoing energy transfer by either donating or accepting excited state energy to the fluorescent ion. The membrane is then exposed to excitation light and the energy transfer between the reagents is measured and related to the membrane potential. The method is useful for detecting changes in membrane potential in subcellular organelle membranes in biological systems. The method is used for screening of test chemicals for activity to modulate the activity of target ion channel. The invention also provides a transgenic organism comprising a first reagent that comprises a charged hydrophobic fluorescent molecule and a second reagent comprising a bioluminescent or naturally fluorescent protein. The present sequence is Discosoma species (red) DNA encoding an Anthozoa fluorescent protein (FP), drFP583

SQ   Sequence 678 BP; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:
```
Pred. No.:
Score:                 2.3e-140
Percent Similarity:    1214.00          Length:           678
Best Local Similarity: 100.00%          Matches:          225
Query Match:           100.00%          Mismatches:       0
DB:                    100.00%          Conservative:     0
                                        Indels:           0
                                        Gaps:             0
```

US-10-006-922A-12 (1-225) x AAD11142 (1-678)

```
Qy      1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAGGA 60

Qy     21 ThrValAlaAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyLysProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGGTCAATGGGCACGAGTTTATTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACAATACCGTAAAGCTTAAGGTTAACCAAGGGGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420
```

Db 361 AAGGTCAAGTTCATTGGGCGTGAACTTTCCTTCCGAATGGAACCTGTTATGCAAAAGAAGACA

Qy 141 MetGlyTrpGluLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db 421 ATGGGCTGGGAAGCCAGCACTGAGCCGTTTGTAATCCTGACGATGGGCGTTGAAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLeuLysSerGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db 481 ATTCATAAAGGCTCTGAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluIleAsnGlyAsnGlyArg 220

Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGATTAATGGAAATGGAAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225

Db 661 CACCACTGTTCCTT 675

RESULT 4

ID ABA95905 standard; cDNA; 678 BP.
XX
AC ABA95905;
XX
DT 29-MAY-2002 (first entry)
XX
DE Yeast optimised RPP encoding cDNA SEQ ID NO 1.
XX
KW Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW Escherichia coli; green fluorescent protein; biotechnology; gene; ss.
XX
OS Anthozoa.
XX
FH Key Location/Qualifiers
FT CDS 1..678
FT /tag= a
FT /product= "yeast optimised Red Fluorescent Protein"
XX
DR DE2001395-U1.
PR 27-JAN-2000; 2000DE-02001395.
PF 27-JAN-2000; 2000DE-02001395.
PD 15-MAR-2001.
PN DE2001395-U1.
XX
PA (GPCB-) GPC BIOTECH AG.
XX
DR WPI; 2002-228394/29.
DR P-PSDB; ABB08834.
XX
PT New DNA encoding red fluorescent protein, useful as marker in
PT biotechnology, has sequence optimised for expression in eukaryotes,
PT especially yeast or plants.
XX
PS Claim 1; Fig 1; 19pp; German.
XX
CC The invention relates to DNA (I) containing either sequence ABA95905 or
CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC (yRFP). (I) are used to express red fluorescent protein (RFP),
CC especially Saccharomyces cerevisiae and
CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC plants, especially dicotyledonous plants including Nicotiana tabacum or
CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC especially Escherichia coli. RFP is useful in the same way as green
CC fluorescent protein but is more generally applicable in modern
CC biotechnology. (I) are optimised for expression in yeast and so generate
CC RFP at higher levels with stronger fluorescence and thus lowers the
CC detection limit and gives a better signal-to-noise ratio

XX
SQ Sequence 678 BP; 198 A; 147 C; 159 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.3e-140 Length: 678
   1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x ABA95905 (1-678)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGATCTTCAAGAACGTCATCAAGGAATTCATGAGATTCAAGGTTAGAATGGAAGGT 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db 61 ...

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAACACTGTCAAGTTGAAGGTTACTAAGGGTGGTCCATTGCCATTCGCTTGGGACATC 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCTCCACAATTCCAATACGGTTCCAAGGTCTACGTCAAGCACCCAGCTGACATTCCA 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAAGTTGTCCTTCCCAGAAGGTTTCAAGTGGGAAAGAGTCATGAACTTCGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGTGTTGTTACTGTTACTCAAGACTCCTCCTTGCAAGACGGTTGTTTCATCTAC 360

Qy 121 LysValLeuGlyPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCATCGGTGTCAACTTCCCATCTGACGGTCCAGTCATGCAAAAGAAGACT 420

Qy 141 MetGlyTrpGluLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGTTGGGAAGCTTCTACCGAAGCGTTTGTACCCAAGAGACGGCGTTCTTGAAGGGTGAA 480

Qy 161 IleHisLysAlaLeuLeuLysSerGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCACAAGGCCTTGAAGTTGGAAGGTCGGTCACTACTGGTCGAATTCAAGTCTATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCTAAGAAGCCAGTCCAATTGCCAGGTTACTACTACGTTGACTCTAAGTTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluIleAsnGlyAsnGlyArg 220
Db 601 ATCACCTCTCACAACGAAGACTACACTATCGTCGAAATCAATGGTAACGGTAGA 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACTGTTCTTG 675

RESULT 5

ID ABA95922 standard; DNA; 678 BP.
XX
AC ABA95922;
XX
DT 29-MAY-2002 (first entry)
XX
DE Yeast optimised RPP related DNA SEQ ID NO 19.
XX
KW Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;

```
KW    Escherichia coli; green fluorescent protein; biotechnology; gene; ss.
XX
OS    Anthozoa.
XX
FH    Key              Location/Qualifiers
FT    CDS              1..678
FT                     /*tag= a
FT                     /product= "yeast optimised Red Fluorescent Protein"
XX
PN    DE20001395-U1.
XX
PD    15-MAR-2001.
XX
PF    27-JAN-2000; 2000DE-02001395.
XX
PR    27-JAN-2000; 2000DE-02001395.
XX
PA    (GPCB-) GPC BIOTECH AG.
XX
DR    WPI; 2002-228394/29.
XX
PT    New DNA encoding red fluorescent protein, useful as marker in
PT    biotechnology, has sequence optimized for expression in eukaryotes,
PT    especially yeast or plants.
XX
PS    Disclosure; Page 14-15; 19pp; German.
XX
CC    The invention relates to DNA (I) containing either sequence ABA95905 or
CC    sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC    (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC    eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC    plants, especially dicotyledonous plants including Nicotiana tabacum or
CC    Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC    especially Escherichia coli. RFP is useful in the same way as green
CC    fluorescent protein but is more generally applicable in modern
CC    biotechnology. (I) are optimised for expression in yeast and so generate
CC    RFP at higher levels with stronger fluorescence and thus lowers the
CC    detection limit and gives a better signal-to-noise ratio. The present
CC    sequence is that of a polynucleotide encoding the yeast optimised RFP,
CC    useful to the invention
XX
SQ    Sequence 678 BP; 202 A; 118 C; 159 G; 199 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:

Score:	2.3e-140	Length:	678
Percent Similarity:	1214.00	Matches:	225
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Conservative:	0
	100.00%	Indels:	0
	6	Gaps:	0

DB:

US-10-006-922A-12 (1-225) x ABA95922 (1-678)

```
Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGATCTTCCAAGAACGTTATTAAGGAATTCATGAGATTCAAGGTTAGAATGGAAGGT 60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACTGTTAACGGTCACGAATTCGAAATTGAAGGTGAAGGTGAAGGTAGACCATACGAAGGT 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACAACACTGTTAAGTTGAAGGTTACTAAGGGTGGTCCATTGCCATTCGCTTGGGACATT 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTGTCTCCACAATTCCAATACGGTTCTAAGGTTTACGTTAAGCACCCAGCTGACATTCCA 240

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACTACAAGAAGTTGTCTTTCCCAGAAGGTTTCAAGTGGGAAAGAGTTATGAACTTCGAA 300
```

```
Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGTGGTGTTGTTACTGTTACTCAAGACTCTTCTTTGCAAGACGGTTGTTTCATTTAC 360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGTTAAGTTCATTGGTGTTAACTTCCCATCTGACGGTCCAGTTATGCAAAAGAAGACT 420

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGGTTGGGAAGCTTCTACTGAAAGATTGTACCCAAGAGACGGTGTTTTGAAGGGTGAA 480

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluTyrSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATTCACAAGGCTTTGAAGTTGAAGGACGGTGGTCACTACTTGGTTGAATATTCAATT 540

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACATGGCTAAGAAGCCAGTTCAATTGCCAGGTTACTACTACGTTGACTCTAAGTTGGAC 600

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATTACTTCTCACAACGAAGACTACACTATTGTTGAAATAACTGAAGGTAGA 660

Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db    661 CACCACTTGTTCTTG 675
```

```
RESULT 6
ABA95921

ID   ABA95921 standard; DNA; 678 BP.
XX
AC   ABA95921;
XX
DT   29-MAY-2002 (first entry)
XX
DE   Yeast optimised RFP related DNA SEQ ID NO 18.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology; gene; ss.
XX
OS   Anthozoa.
XX
FH   Key              Location/Qualifiers
FT   CDS              1..678
FT                    /*tag= a
FT                    /product= "yeast optimised Red Fluorescent Protein"
XX
PN   DE20001395-U1.
XX
PD   15-MAR-2001.
XX
PF   27-JAN-2000; 2000DE-02001395.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
DR   WPI; 2002-228394/29.
DR   P-PSDB; ABB08834.
XX
PT   New DNA encoding red fluorescent protein, useful as marker in
PT   biotechnology, has sequence optimized for expression in eukaryotes,
PT   especially yeast or plants.
XX
PS   Disclosure; Page 14; 19pp; German.
XX
CC   The invention relates to DNA (I) containing either sequence ABA95905 or
CC   sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC   plants, especially dicotyledonous plants including Nicotiana tabacum or
```

CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern
CC   biotechnology. (I) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio. The present
CC   sequence is that of a polynucleotide encoding the yeast optimised RFP,
CC   useful to the invention
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.:
Score: 2.3e-140
                         1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
                         6   Indels: 0
                                        Gaps: 0
XX

US-10-006-922A-12 (1-225) x ABA95921 (1-678)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluProValIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTGAAATAGAAGGCGAAGGCCATAGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGluGlyProLeuGluTrpGluValMetAsnGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGAGTTAAATGGGAAAGGGTCATGAACTTTGGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGANTCCAGTTTGCANGANGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGANGGACCTGTTATGCANAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGGGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluGlyThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221 HisHisHisLeuPheLeu 225
         ||||||||||||
Db  661 CACCATCTGTTCCTT 675
```

RESULT 7
ID   ABA95920 standard; RNA; 678 BP.
XX

AC   ABA95920;
XX
DT   29-MAY-2002 (first entry)
XX
DE   Yeast optimised RFP encoding RNA SEQ ID NO 16.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology; ss.
XX
OS   Anthozoa.
XX
PN   DE20001395-U1.
XX
PD   15-MAR-2001.
XX
PF   27-JAN-2000; 2000DE-02001395.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PS   Disclosure; Page 13; 19pp; German.
XX
DR   WPI; 2002-228394/29.
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
PT   New DNA encoding red fluorescent protein, useful as marker in
PT   biotechnology, has sequence optimized for expression in eukaryotes,
PT   especially yeast or plants.
XX
CC   The invention relates to DNA (II) containing either sequence ABA95905 or
CC   sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially yeast, especially Saccharomyces cerevisiae in
CC   plants, especially dicotyledonous plants including Nicotiana tabacum or
CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern
CC   biotechnology. (I) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio. The present
CC   sequence is that of an RNA sequence corresponding to the yeast optimised
CC   RFP encoding cDNA given as SEQ ID NO 1 (ABA95905)
XX
SQ   Sequence 678 BP; 198 A; 147 C; 159 G; 0 T; 174 U; 0 Other;
XX

Alignment Scores:
Pred. No.:
Score: 2.3e-140
                         1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
                         6   Indels: 0
                                        Gaps: 0
XX

US-10-006-922A-12 (1-225) x ABA95920 (1-678)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AUGAGGUCUUCCAAGAAUCGUUAUCAAGGAGUUCAUGAGGUUCAAGGUUCGCAUGGAAGGU 60

Qy   21 ThrValAsnGlyHisGluProValIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACUGUUAACGGUCACGAAUUCGAAAUCGAAGGUGAAGGUGAAGGACCAUACGAAGGU 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAACACUGUCAAGCUUAAGGUAACGAAGGUGGUCCAUUGCCAUUCGCUUGGGACAUC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 UUGUCUCCACAAUUCCAAUACGGUUCUAAGGUCUACGUCAAGCACCCAGCUGACAUUCCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGluGlyProLeuGluTrpGluValMetAsnGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ...
```

Db   241 GACUACAAGAAGUGUGUCCUCCAGAAGAAGUGGGAAAGAGAGUCAUGAACUCGAA
|||||||||||||||||||||||
Qy   241 ...

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspLysCysPheIleTyr 120

Db   301 GACGGUGGUGUGUACUGUACUCGAAGACUCCCUUCAGAGACGGUUGUUUCAUCUAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db   361 AAGGUCAAGUUCAUCGGUGUCAACUUCCCAUCUGACGGUCCACUCAUGCAAAAGAAGACU 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db   421 AUGGGUUGGGAAGCUUCUACCGAACGUUUGUACCCAAGAGACGGUGUCUUGAAGGGUGAA 480

Qy   161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db   481 AUCCACAAGGCCUUGAAGUUGAAGGACGGUGGUCACUACUUGGUGGAAUUCAAGUCUAUC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Db   541 UACAUGGCUAAGAAGCCAGUCCAAUUGCCAGGUUACUACUACGUUGACUCUAAGUUGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220

Db   601 AUCACCUCUCACAACGAAGACUACACUAUCGUCGAACAAUACGAACGUACUGAAGGUAGA 660

Qy   221 HisHisLeuPheLeu 225
     |||||||||||||||
Db   661 CACCACUUGUUCCUG 675

RESULT 8

ID   AAD46278

AC   AAD46278;

DT   27-DEC-2002 (first entry)

DE   Discosoma sp. drFP583 (NFP-6) wild-type protein encoding DNA.

KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; gene; ds.

OS   Discosoma sp.

FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /tag= a
FT   /product= "drFP583 wild-type protein"

XX
PN   WO200268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-0006922.
XX
PA   (CLON-) CLONTECH LAB INC.
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   P-PSDB; AAE28833.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page 70; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is a DNA encoding Discosoma sp. drFP583 (NFP-6) wild-type
CC   protein of the invention
XX
SQ   Sequence 678 BP; 146 A; 223 C; 203 G; 106 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AAD46278 (1-678)

Alignment Scores:
Pred. No.:   1
Score: 2.3e-140
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Length: 678
Query Match: 100.00%   Matches: 225
DB: 6   Conservative: 0
     Mismatches: 0
     Indels: 0
     Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGCGCTCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   481 ATCCACCACGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy   221 HisHisLeuPheLeu 225
     |||||||||||||||

RESULT 9

Db   661 CACCACCTGTTCCTG 675

AC   AAD28207;
XX
DT   22-APR-2002 (first entry)
XX
AC   AAD28207;
XX
DT   AAD28207 standard; DNA; 678 BP.
XX
ID   AAD28207 standard; DNA; 678 BP.
XX
KW   fluorescent timer protein; protein movement; transocation; trafficking;
KW   promoter activity; gene expression; transgenic plant; gene modification;
KW   protein age; anthozoa protein; drFp583; ds.
XX
OS   Discosoma sp.
XX
DE   Discosoma sp. humanised wild-type Anthozoa protein drFP583 DNA.
XX
FH   Key   Location/Qualifiers
FT
FT   CDS   1..678
FT   /tag= a
FT   /product= "Humanised wild-type Anthozoa protein drFP583"
XX
PN   WO200196373-A2.
XX
PD   20-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-US019097.
XX
PR   14-JUN-2000; 2000US-0211607P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Fradkov AF, Terskikh A;
XX
DR   WPI; 2002-154595/20.
DR   P-PSDB; AAE17540.
XX
PT   New fluorescent timer proteins comprising an emission spectrum that
PT   changes over time from a first wavelength to a second wavelength, useful
PT   for monitoring intracellular protein movement, translocation, trafficking
PT   or stability.
XX
PS   Example 1; Fig 1; 89pp; English.
XX
CC   The invention relates to a fluorescent timer protein having an emission
CC   spectrum that changes over time after synthesis from a first wavelength
CC   to a second wavelength. The fluorescent timer proteins are useful in
CC   monitoring the activity of a promoter, determining the age of a protein,
CC   identifying an agent that modulates the activity of a promoter and in
CC   enriching a population of cells comprising a fluorescent timer protein.
CC   The fluorescent timer proteins are also useful for assessing gene
CC   expression during development of a multicellular organism or during
CC   cellular differentiation, in response to a drug or other inducer of
CC   promoter activity, as a reporter to serve as a read-out of promoter
CC   activity, monitoring intracellular protein movement or translocation,
CC   protein trafficking, or protein stability, to investigate temporal
CC   aspects of the activity of a regulatory element, for determining cell
CC   fate during development and organ remodelling, in spatial and temporal
CC   visualisation of newly synthesised proteins and accumulated proteins, and
CC   in distinguishing between newly formed and pre-existing structures, e.g.
CC   membrane junctions and extracellular matrix components. The fluorescent
CC   timer proteins may further be used to investigations where photobleaching
CC   techniques are employed, as detectable labels, as selectable markers, as
CC   biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC   assays, and as second messenger detectors. The nucleic acids can be used
CC   to generate transgenic, non-human plants or animals or site-specific gene
CC   modifications in cell lines. The present sequence is a DNA encoding
CC   Discosoma sp. humanised wild-type Anthozoa protein drFP583 used for
CC   generating fluorescent proteins
XX
SQ   Sequence 678 BP; 146 A; 223 C; 203 G; 106 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.3e-140   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 100.00%   Indel8:   0
   6   Gaps: 0

US-10-006-922a-12 (1-225) x AAD28207 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCGCTTCCCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy   221 HisHisLeuPheLeu 225
       |||||||||||||||
Db   661 CACCACCTGTTCCTG 675

RESULT 10

AC   ADC24127;
XX
ID   ADC24127;
XX
AC   ADC24127 standard; DNA; 678 BP.
XX
DT   18-DEC-2003 (first entry)
XX
DE   Discosoma wild-type red fluorescent protein DNA #1.
XX
KW   Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW   fluorescent protein variant; transcription induction detection;
KW   fluorescence energy resonance transfer; FRET; protein kinase;
KW   protein phosphatase; ion indicator; ds.
XX

OS   Discosoma.
XX
PN   US2003059835-A1.
XX
PD   27-MAR-2003.
XX
PF   10-APR-2002; 2002US-00121258.
XX
PR   26-FEB-2001; 2001US-00794308.
PR   24-MAY-2001; 2001US-00865538.
XX
PA   (TSIE/) TSIEN R Y.
PA   (CAMP/) CAMPBELL R E.
XX
PI   Tsien RY, Campbell RE;
XX
DR   WPI; 2003-743764/70.
DR   P-PSDB; ADC24126.
XX
PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.
XX
PS   Example 1; SEQ ID NO 2; 67pp; English.
XX
CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   conaining a vector which comprises (I) operatively linked to an
CC   expression control sequence, and an unit to assay the variant fluorescent
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescence protein produced by (VII), where variant fluorescent protein
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcription, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This sequence encodes Discosoma wild-
CC   type red fluorescent protein.
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADC24127 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheValValAsnGlyGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Alignment Scores:
Pred. No.:
Score: 2.3e-140
                       1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 10

Length: 678
Matches: 225
Mismatches: 0
Conservative: 0
Indels: 0
Gaps: 0

RESULT 11
ADF70404
ID   ADF70404 standard; cDNA; 678 BP.
XX
AC   ADF70404;
XX
DT   12-FEB-2004 (first entry)
XX
DE   Discosoma wild-type GFP variant cDNA SeqID27.
XX
KW   ligand; orphan receptor protein; fusion protein; fluorescent protein;
KW   cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
KW   GFPuv; Enhanced GFP; EGFP; gene; ss.
XX
OS   Discosoma sp.
XX
PN   WO2003071272-A1.
XX
PD   28-AUG-2003.
XX
PF   21-FEB-2003; 2003WO-JP001901.
XX
PR   22-FEB-2002; 2002JP-00045728.
PR   23-JUL-2002; 2002JP-00213949.
PR   11-OCT-2002; 2002JP-00298237.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX
DR   WPI; 2003-697654/66.

Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAGGGCGAAGGAGAGGGCCATACGGAGGC 120

Qy   41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTTAACCAAGGGGGGCCGCTTCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTCAGTATGGAAGCAAGGTTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTTGTTACTGTGACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAAGGGCGC 660

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCATCTGTTCCTT 675

Left column

```
DR     P-PSDB; ADF70403.
XX
PT     Transformation of cells with a fusion protein of an orphan receptor
PT     protein with a fluorescent protein useful for identification of ligands
PT     to the orphan receptor.
XX
PS
XX
CC     Disclosure; SEQ ID NO 27; 594pp; Japanese.
XX
CC     This invention relates to a novel method of identifying ligands to an
CC     orphan receptor protein which comprises transforming cells with DNA
CC     encoding a fusion protein of the orphan receptor with a fluorescent
CC     protein, so that the fusion protein is expressed in the cells (or cell
CC     membranes isolated from them) and contacting the cells with the potential
CC     ligand to be tested. A suitable fluorescent protein for incorporation in
CC     the fusion protein is green fluorescent protein (GFP), for example GFP-1,
CC     wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
CC     identification of ligands binding to an orphan receptor protein.
XX
SQ     Sequence 678 BP; 204 A; 129 C; 179 G; 166 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.:
Score:                  2.3e-140
                        1214.00
Percent Similarity:     100.00%        Length:        678
Percent Identity:       100.00%        Matches:       225
Best Local Similarity:  100.00%        Conservative:  0
Query Match:            10             Mismatches:    0
DB:                                    Indels:        0
                                       Gaps:          0

US-10-006-922A-12 (1-225) x ADF70404 (1-678)

Qy    1  MetArgSerLysGluAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAAGATGTTATCAAGGAGTTTATGAGGTTTAAGGTTCGTGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGATACAGAGGC  120

Qy   41  HisAsnThrHisLysLeuLysIleThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCCATAAAGCTTAAGGTAACCAAGGGAGGACCTTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy   81  AspThrTyrLysLeuSerPheProGluGlyTyrProGluValArgValMetAsnPheGlu  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACACATATAAAAGCTCTCATTTCCTGAAGGAATTTAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATAAGGCCCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660
```

Right column

```
Qy  221  HisHisLeuPheLeu  225
         |||||||||||||||
Db  661  CACCACCTGTTCCTT  675

RESULT 12

ID    ADL46204 standard; DNA; 678 BP.
XX
AC    ADL46204;
XX
DT    20-MAY-2004 (first entry)
XX
DE    Discosoma red fluorescent protein (DsRed) coding sequence.
XX
KW    dsj; gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
KW    oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS    Discosoma sp.
XX

FH    Key             Location/Qualifiers
FT    CDS             1..678
FT                    /*tag= a
FT                    /product= "DsRed protein"
XX
PN    WO2003086446-A1.
XX
PD    23-OCT-2003.
XX
PF    09-APR-2003; 2003WO-US010879.
XX
PR    10-APR-2002; 2002US-00121258.
PR    29-JUL-2002; 2002US-00209208.
XX
PA    (REGC ) UNIV CALIFORNIA.
XX
PI    Tsien RY, Campbell RE, Baird GS;
XX
DR    WPI; 2003-845265/78.
DR    P-PSDB; ADL46203.
XX
PT    New monomeric and dimeric Anthozoan fluorescent protein variants with
PT    reduced propensity to oligomerize, and encoding polynucleotides, useful
PT    in molecular biology, e.g. in immunoassays or in tracking protein
PT    movement in cells.
XX
PS    Disclosure; SEQ ID NO 2; 166pp; English.
XX
CC    The invention relates to a polynucleotide sequence encoding a Discosoma
CC    red fluorescent protein (DsRed) variant having a reduced propensity to
CC    oligomerize. The protein variant comprises one or more amino acid
CC    substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC    sequence, where the substitutions result in reduced propensity of the
CC    DsRed variant to form tetramers and where the variant displays detectable
CC    fluorescence of at least one red wavelength. The composition and methods
CC    are useful in producing red fluorescent proteins having reduced
CC    propensity for oligomerization, especially tetramerization. The protein
CC    may be used in molecular biology and in other scientific applications,
CC    such as in immunoassays or hybridization assays, or in tracking the
CC    movement of proteins in cells. This sequence corresponds to the DsRed
CC    coding sequence.
XX
SQ    Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.:
Score:                  2.3e-140
                        1214.00
Percent Similarity:     100.00%        Length:        678
Percent Identity:       100.00%        Matches:       225
Best Local Similarity:  100.00%        Conservative:  0
Query Match:            11             Mismatches:    0
DB:                                    Indels:        0
                                       Gaps:          0

US-10-006-922A-12 (1-225) x ADL46204 (1-678)
```

```
Qy    1  MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGGATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC    120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGGAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221  HisHisHisLeuPheLeu    225
         |||||||||||||||||
Db  661  CACCATCTGTTCCTT      675
```

```
RESULT 13
ADN33978

ID   ADN33978 standard; DNA; 678 BP.
XX
AC   ADN33978;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Wild-type DsRED encoding sequence.
XX
KW   da.
OS   Discosoma sp.
OS   Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED;
XX
FF   18-DEC-2002; 2002WO-US040539.
PN   WO200305415B-A2.
PD   03-JUL-2003.
PR   19-DEC-2001; 2001US-0341723P.
XX
PA   (UYCH-) UNIV CHICAGO.
```

```
XX
PI   Bevis B, Glick B;
XX
DR   WPI; 2003-569236/53.
DR   P-PSDB; ADN33979.
XX

Claim 8; SEQ ID NO 1; 65pp; English.

The present invention relates to nucleic acid that encodes a rapidly
maturing chromo or fluorescent mutant of a Cnidarian chromo- or
fluorescent protein or its mutant. The protein is useful in applications
involving nucleic acid encoding a chromo- or fluorescent protein and is
useful for producing a chromo and/or fluorescent protein which involves
growing the cell, whereby the protein is expressed, and isolating the
protein substantially free of other proteins. The protein is useful in
applications involving chromo- or fluorescent protein and is useful as
PCR primers, hybridization probes, etc. The expression cassettes are
useful for synthesizing related proteins. The chromoproteins are useful
as coloring agents which are capable of imparting color or pigment to a
particular composition of matter e.g. food compositions, pharmaceuticals,
cosmetics, living organisms, e.g., animals and plants. The chromoproteins
may also find use as labels in analyte detection assays, e.g. in
biological analyses of interest and as selectable markers in recombinant
DNA applications, e.g. the production of transgenic cells and organisms.
The fluorescent proteins find use in a variety of different applications,
e.g. in fluorescence resonance energy transfer (FRET) applications, as
biosensors in prokaryotic and eukaryotic cells, in applications involving
the automated screening of arrays of cells expressing fluorescent
reporting groups by using microscopic imaging and electronic analysis, as
second messenger detectors, and in fluorescence activated cell sorting
applications and as in vivo marker in animals. The fluorescent proteins
also find use in protease cleavage assays. The proteins can also be used
is assays to determine the phospholipid composition in biological
membranes and as a fluorescent timer. The present sequence represents the
wild-type DsRED encoding sequence.
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                                  2.3e-140
Score:                                     1214.00
Percent Similarity:                        100.00%
Best Local Similarity:                     100.00%
Query Match:                                 11

                        Length:   678
                        Matches:  225
                        Conservative:  0
                        Mismatches:    0
                        Indels:        0
                        Gaps:          0
```

```
US-10-006-922A-12 (1-225) x ADN33978 (1-678)
```

```
Qy    1  MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGGATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC    120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGGAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
```

Left column

Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCCTGTTTCATCTAC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGGGTGGTGTTGAAAGGAGAG 480

Qy 161 IleHisSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlySerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATTCACATAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCATCTGTTCCTT 675

OS Discosoma sp.

KW target detection; gene; red fluorescent protein; dsI; RED.
KW fluorobody; binding ligand; green fluorescent protein; GFP;
DE Discosoma sp. red fluorescent protein (RED) DNA.
DT 22-APR-2004 (first entry)
DT 22-APR-2002; 2002US-00132067.
AC ADI36420;
XX
ID ADI36420 standard; DNA; 678 BP.
RSULT 14
ADI36420

FH Key   Location/Qualifiers
FT CDS   1..678
FT   /+tag= a
FT   /product= "Discosoma sp. red fluorescent protein (RED)"

PN US20030203355-A1.
PD 30-OCT-2003.
PF 24-APR-2002; 2002US-00132067.
PR 24-APR-2002; 2002US-00132067.
PA (LALA-) LOS ALAMOS NAT LAB.
PA (REGC) UNIV CALIFORNIA.
PI Bradbury AM, Zeytun A, Waldo GS;
XX
PT Novel binding ligand with intrinsic fluorescence and comprising
PT fluorescent protein having heterologous binding sites, useful for
PT detecting target molecule.
XX
DR WPI; 2004-154325/15.
DR P-PSDB; ADI36421.
XX
PS Example 6; SEQ ID NO 3; 23pp; English.
XX
CC The invention relates to binding ligands (fluorobodies) with intrinsic
CC fluorescence, which comprises green fluorescent protein (GFP) having
CC heterologous binding sites. The binding ligand is useful for detecting
CC the target molecule and is efficiently detects the target molecule. The

Right column

CC present sequence is Discosoma sp. red fluorescent protein (RBD) DNA used
CC in the exemplification of the invention.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 12   Mismatches: 0
                                        Indels: 0
                                        Gaps: 0

DB:

US-10-006-922A-12 (1-225) x ADI36420 (1-678)

Qy   1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGGTTTAAATGGGAAAGGGTTATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGGGTGGTGTTGAAAGGAGAG 480

Qy 161 IleHisSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlySerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATTCACATAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCATCTGTTCCTT 675

RSULT 15
ADM97768
ID ADM97768 standard; DNA; 678 BP.
AC ADM97768;
XX
DT 01-JUL-2004 (first entry)
XX
DE D sp red fluorescent protein coding sequence SEQ ID NO: 21.
XX

KW d8; gene; enzyme; sensor cell; fluorescent protein;
KW signal transduction detection system; promoter; targeting sequence;
KW targeted drug.
XX
OS Discosoma sp.
XX
FH Key   Location/Qualifiers
FT CDS   1..678
FT   /*tag= a
FT   /product= "fluorescent protein"
XX
PN WO200403141S-A2.
XX
PD 15-APR-2004.
XX
PF 05-SEP-2003; 2003WO-US028078.
PR 05-SEP-2002; 2002US-0408297P.
XX
PA (VERT-) VERTEX PHARM INC.
XX
PI Whitney MA, Zeh K, Sanders PS;
XX
DR WPI; 2004-330208/30.
DR P-PSDB; ADM97769.
XX
PT Developing a sensor cell, useful in determining the activity of a target
PT gene and in developing therapeutic drugs, comprises providing cells
PT comprising a signal transduction detection system and introducing DNA
PT construct into cells.
XX
PS Disclosure; Page 167-168; 234pp; English.
XX
CC The present invention relates to a method of developing a sensor cell,
CC for determining the activity of a target gene in the cell, which
CC comprises providing a homogeneous population of cells, where each of the
CC cells comprises a signal transduction detection system and introducing
CC into the population of cells an isolated DNA construct comprising a
CC promoter operatively linked to a targeting sequence. The method is useful
CC in developing a sensor cell for determining the activity of a target gene
CC in the cell. The sensor cell and the methods are useful in developing new
CC and therapeutic drugs directed to the targets. The present sequence is a
CC coding sequence shown in the exemplification of the invention.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Db   241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTCATGAAACTTTGAA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1 AspGlyGlyGlyValValThrValThrGlnAsnSerSerLeuGlnAspGlyCysPheIleTyr   20
Db   1 ATGAGGGTCTTCCAAGAANTGTTATCAAGGAGTTCATGAGGTTTCGCATGGGAAGGA   60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db   61 ACGGTCAANGGGCACGAGTTTGAAATAGAAGGCGAAGGCGAGGGCCATACGAAGGC   120

Qy   41 HisAsnThrValLysLeuLysLeuValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db   121 CACAATACCGTAAAGCTTAAGGTAACCAAAGGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db   181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100

Qy   101 AspGlyGlyValValThrValThrGlnAsnSerSerLeuGlnAspGlyCysPheIleTyr   120
Db   301 GACGGTGGCGTCGTTACTGTTACCCAGAATTCCAGTTTGCAGGATGGCTGTTTCATCTAC   360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
Db   361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA   420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
Db   421 ATGGGCTGGGAAGCCAGCACACTGAGCGGTTTGTATCCTCGTGATGGCGTGTTGAAAGGGGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
Db   481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT   540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
Db   541 TACATGGCCAAAGAAGCCTTGCCAGGCTACTACTATGTTGACTCCAAAACTGGAT   600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
Db   601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGGAGGC 660

Qy   221 HisHisLeuPheLeu 225
Db   661 CACCATCTGTTCCTT 675

XX
CC
US-10-006-922A-12 (1-225) x ADM97768 (1-678)

Alignment Scores:
Pred. No.:
Score: 2.3e-140
Score: 1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity:100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 12   Indels: 0
                                      Gaps: 0

Search completed: January 12, 2006, 10:16:26
Job time : 485 secs

THIS PAGE BLANK (USPTO)

OM protein - nucleic search, using frame_plus_p2n model

```
                        Gencore version 5.1.6
                        Copyright (c) 1993 - 2006 Compugen Ltd.

Run on:      January 12, 2006, 09:14:42 ; Search time 3744 Seconds
                                           (without alignments)
                                           2811.724 Million cell updates/sec

Title:       US-10-006-922A-12
Perfect Score: 1214
Sequence:    1 MRSSKNVIKEFMRFKVRMEG..........EDYTIVEQYERTEGRHHLFL 225

Searched:    41078325 seqs, 2339541228 residues

Total number of hits satisfying chosen parameters:    82156650

Scoring table: BLOSUM62
             Xgapop 10.0 , Xgapext 0.5
             Ygapop 10.0 , Ygapext 0.5
             Fgapop  6.0 , Fgapext 7.0
             Delop   6.0 , Delext  7.0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat 10012006 162408 18933/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN 1.1.8010 @runat_10012006_162408_18933 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
             1: gb_est1:*
             2: gb_est2:*
             3: gb_est3:*
             4: gb_htc:*
             5: gb_est4:*
             6: gb_est5:*
             7: gb_est6:*
             8: gb_est7:*
             9: gb_gss1:*
            10: gb_gss2:*
            11: gb_gss3:*
```

SUMMARIES

No.	Score	%	Len	DB	ID	Description
8	216.5	17.8	606	5	BW741633	BW741633 BI379468 BFLG1_001
9	216.5	17.8	637	3	BI379468	BI379468 BFLG1_002
10	216.5	17.8	687	3	BI387917	BI387917 BFL26_001
11	212.5	17.5	643	3	BI386699	BI386699 BFLG2_000
12	212.5	17.5	675	3	BI382638	BI382638
13	210.5	17.3	590	5	BW883955	BW883955 BW883955
14	210.5	17.3	812	5	BW893733	BW893733
15	209.5	17.3	790	1	AL044652	AL044652 DXF2p434P
16	206.5	17.0	630	5	BW731527	BW731527 BW731527
17	206.5	17.0	677	5	BW894392	BW894392 BW894392
18	206.5	16.9	605	5	BW744973	BW744973 BW744973
19	205.5	16.9	618	5	BW871203	BW871203 BW871203
20	205.5	16.9	668	5	BW894220	BW894220 BW894220
21	205.5	16.8	602	5	BW845436	BW845436 BW845436
22	203.5	16.8	602	5	BW872238	BW872238 BW872238
23	203.5	16.8	620	5	BW870345	BW870345 BW870345
24	203.5	16.8	630	5	BW740938	BW740938 BW740938
25	203.5	16.8	641	5	BW798018	BW798018 BW798018
26	203.5	16.8	641	5	BW881726	BW881726 BW881726
27	203.5	16.8	647	5	BW736374	BW736374 BW736374
28	203.5	16.8	652	5	BW733673	BW733673 BW733673
29	203.5	16.8	652	5	BW881040	BW881040 BW881040
30	202.5	16.8	652	5	BW870882	BW870882 BW870882
31	202.5	16.7	744	8	DR622961	DR622961 EST101308
32	202.5	16.7	573	5	BW865453	BW865453 BW865453
33	201.5	16.6	675	5	BW895120	BW895120 BW895120
34	201.5	16.6	722	8	DR624559	DR624559 EST101468
35	200.5	16.5	730	8	DR623458	DR623458 EST101358
36	200.5	16.5	736	8	DR629310	DR629310 EST101943
37	200.5	16.5	768	8	DR629433	DR629433 EST101956
38	200.5	16.5	791	8	DR621827	DR621827 EST101195
39	200.5	16.5	795	8	DR624841	DR624841 EST101496
40	200.5	16.5	818	8	DR627300	DR627300 EST101742
41	200.5	16.5	828	8	DR622639	DR622639 EST101276
42	200.5	16.5	831	8	DR628628	DR628628 EST101875
43	200.5	16.5	833	8	DR624844	DR624844 EST101497
44	200.5	16.5	840	8	DR624776	DR624776 EST101490
45	200.5	16.5	862	8	DR624009	DR624009 EST101413

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	461	38.0	556	8	DR987865	DR987865 JGI_AOSF1
2	461	38.0	556	8	DR987866	DR987866 JGI_AOSF1
3	225	18.5	556	3	BI385436	BI385436 BFL26_000
4	218.5	18.0	508	5	BW802012	BW802012 BW802012
5	218.5	18.0	646	5	BW736869	BW736869 BW736869
6	217.5	17.9	639	5	BW736108	BW736108 BW736108
7	216.5	17.8	604	5	BW738638	BW738638 BW738638

RESULT 1
DR987865/c

```
LOCUS       DR987865          556 bp    mRNA       linear    EST 03-AUG-2005
DEFINITION  JGI_AOSF1131.rev AOSF Montastraea faveolata adult colony
            Montastraea faveolata cDNA clone AOSF1131 3', mRNA sequence.
ACCESSION   DR987865
VERSION     DR987865.1  GI:71782571
KEYWORDS    EST.
SOURCE      Montastraea faveolata
  ORGANISM  Montastraea faveolata
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Faviina; Faviidae; Montastraea.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Schwarz,J.A., Brokstein,P., Manohar,C., Coffroth,M.A., Szmant,A.
            and Medina,M.
  TITLE     Coral-Symbiodinium EST Project
  JOURNAL   Unpublished (2005)
COMMENT     Other ESTs: JGI_AOSF1131.fwd
            Contact: Schwarz, JA, Medina, M.
            Evolutionary Genomics
            DOE Joint Genome Institute
            2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            Tel: 925-296-5823
            Email: jaschwarz@lbl.gov
            http://www.jgi.doe.gov
            cDNA Library Preparation: DOE Joint Genome Institute:
            http://www.jgi.doe.gov
            DNA Sequencing: DOE Joint Genome Institute:
            Naming Conventions: EST name is generated by the concatenation of
            the JGI Clone Id and the direction of sequencing. The suffix '.rev'
            indicates a reverse sequencing read of the insert. It does not
            necessarily reflect the orientation of the insert.
```

ALIGNMENTS

Poly-T: A run of 14 or more T residues at the beginning of this sequence has been removed. Small Insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases. Plate: AOSF 0009 row: f column: 20
High quality sequence stop: 556.

FEATURES   Location/Qualifiers
     source   1..556
                     /organism="Montastraea faveolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:48498"
                     /clone="AOSF1131"
                     /dev_stage="Adult colony"
                     /lab_host="ElectroMAX DH10B"
                     /note="vector: pDNR-LIB; Site_1: SfiI; Site_2: SfiI; The
                     library was prepared from total RNA using the Creator
                     SMART cDNA Library Construction Kit with the LD-PCR method
                     to amplify the cDNA. Amplified cDNA was digested with
                     SfiI, size selected for >400bp, and ligated into the
                     pDNR-LIB vector. WARNING: this library contains a small
                     percentage of cDNAs derived from the coral's symbiont,
                     Symbiodinium sp."

ORIGIN

US-10-006-922A-12 (1-225) x DR987865

Alignment Scores:
Pred. No.:
Score: 3.93e-45
Percent Similarity: 62.9%   Length: 556
Best Local Similarity: 43.92%   Matches: 83
Query Match: 37.97%   Conservative: 36
                                    Mismatches: 46
                                    Indels: 24
                                    Gaps: 1

Qy   32 GluGlyGlyGlyArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGly 51
            |||||||||||||||| :::|||||||||||||||||||||||||| |||
Db   556 GACGGAGAGGCAAGCCTTTCGAGGGAAAACACAGTATAAACCTTCAAGTCCAAGAAGGT 497

Qy   52 GlyProLeuProPheAlaAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysVal 71
            ||||||||||||||||||:::|||:::
Db   496 GGACCTCTCCCTTTTGCTTACGATATCTTGACGACAGCAATTCTCGTACGGACACAGGGCA 437

Qy   72 TyrValLysHisProAlaAspIleProAspTyrLysLysValValAsnPhePheIleGly 91
            :::||| :::|||||:::|||||
Db   436 TTCACCAAATACCCAAGAGACCTCCCAGACTATTTCAAGCAGACATC 377

Qy   92 LysTrpGluArgValMetAsnPheGluArgProGlyIleGlyValValThrValThrGlnSer 111
            ||||||:::||| ||| |||:::
Db   376 TCCTGGGAACGTGTGTATGACTTTGACCGTGTCAACAACATC 317

Qy   112 SerLeuGlnAspGlyCysPheIleLeuTyrLysValLeuValAlaAsnPheProSer 131
            :::||:::
Db   316 AAAATGGAGGGCGGCGACCGTTTATCTATGATAATTCGATTCATGTGGTTAACTTCCTGCT 257

Qy   132 AspGlyProValMetGlyThrMetGlyTyrTrpGluAlaSerThrGluArgLeuTyr 151
            |||||||||||
Db   256 AATGGTCCAATTATGCAGAAGAGAGACT 230

Qy   152 ProArgAspGlyValLeuLysGlyLeuHisLysAlaLeuLysLysAspGlyGlyGly 171
            ||||||:::
Db   229 ----- CTGTTGCTTGAAGGAAATAAG 209

Qy   172 HisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysProValGlnLeuProGly 191
            ||| :::|||::
Db   208 CATCACCGATGTAACTTCAGAAGTACTTAAAGCCACAGAAGGATGTTGGTGGTTGCCAGAA 149

Qy   192 TyrTyrTyrValAspSerLysLeuAsnGluThrSerHisAsnGluAspTyrThrIleVal 211
            |||:::||| |||
Db   148 TATACACTTTGTTGACCACCGGAATTGAGAATATTAAGTCATGACAAGAGATTACAACAGTG 89

Qy   212 GluGlnTyrGluArgThrGluGlyArg 220
            ||| |||||||||
Db   212 GTGGTCTTATGAGAATGCTGCTGCCCCGC 62

Db
RESULT 2

LOCUS   DR987866   556 bp   mRNA   linear   EST 03-AUG-2005
DEFINITION JGI AOSF1131.fwd AOSF Montastraea faveolata adult colony
            Montastraea faveolata cDNA clone AOSF1131 5', mRNA sequence.
ACCESSION   DR987866
VERSION   DR987866.1 GI:71782572
KEYWORDS   EST.
SOURCE   Montastraea faveolata
  ORGANISM Montastraea faveolata
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Faviina; Faviidae; Montastraea.
REFERENCE   1 (bases 1 to 556)
  AUTHORS   Schwarz,J.A., Brokstein,P., Manohar,C., Coffroth,M.A., Sznant,A.
            and Medina,M.
  TITLE   Coral-Symbiodinium EST Project
  JOURNAL   Unpublished (2005)
COMMENT   Other ESTs: JGI AOSF1131.rev
            Contact: Schwarz, JA, Medina, M.
            Evolutionary Genomics
            DOE Joint Genome Institute
            2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            Tel: 925-296-5823
            Email: jaschwarz@lbl.gov
            cDNA Library Preparation: DOE Joint Genome Institute:
            http://www.jgi.doe.gov
            DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
            Naming Conventions: EST name is generated by the concatenation of
            the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
            indicates a forward sequencing read of the insert. It does not
            necessarily reflect the orientation of the insert.
            Small Insert: Based upon one or more sequencing reads of this clone
            where vector sequence was present at both ends, this clone has been
            determined to contain a cDNA insert on the order of 600-1000 bases.
            Plate: AOSF 0009 row: f column: 20
            High quality sequence stop: 557.

FEATURES   Location/Qualifiers
     source   1..556
                     /organism="Montastraea faveolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:48498"
                     /clone="AOSF1131"
                     /dev_stage="Adult colony"
                     /lab_host="ElectroMAX DH10B"
                     /note="vector: pDNR-LIB; Site_1: SfiI; Site_2: SfiI; The
                     library was prepared from total RNA using the Creator
                     SMART cDNA Library Construction Kit with the LD-PCR method
                     to amplify the cDNA. Amplified cDNA was digested with
                     SfiI, size selected for >400bp, and ligated into the
                     pDNR-LIB vector. WARNING: this library contains a small
                     percentage of cDNAs derived from the coral's symbiont,
                     Symbiodinium sp."

ORIGIN

US-10-006-922A-12 (1-225) x DR987866 (1-556)

Alignment Scores:
Pred. No.:
Score: 3.93e-45
Percent Similarity: 62.9%   Length: 556
Best Local Similarity: 43.92%   Matches: 83
Query Match: 37.97%   Conservative: 36
                                    Mismatches: 46
                                    Indels: 24
                                    Gaps: 1

Qy   32 GluGlyGlyGlyArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGly 51
            |||||||||||||||| :::|||||||||||||||||||||||||| |||
Db   1 GACGGAGAGGCAAGCCTTTCGAGGGAAAACACAGTATAAACCTTCAAGTCCAAGAAGGT 60

Qy   52 GlyProLeuProPheAlaAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysVal 71

```
RESULT 3
BI385436
LOCUS       BI385436    661 bp    mRNA    linear    EST 26-AUG-2003
DEFINITION  MPMGp531) Amphioxus 26hr cDNA library (Name convention: BFL26 or
            MPMGp531) Branchiostoma floridae cDNA clone cDNA clone MPMGp531H0446 5', mRNA
            sequence.
ACCESSION   BI385436
VERSION     BI385436.1  GI:30920419
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1 (bases 1 to 661)
  AUTHORS   Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
            Herwig,R., Vingron,M. and Lehrach,H.
  TITLE     New evidence for genome-wide duplications at the origin of
            vertebrates using an amphioxus gene set and completed animal
            genomes
  JOURNAL   Genome Res. 13 (6A), 1056-1066 (2003)
   PUBMED   12799346
COMMENT     Contact: Panopoulou G
            Laboratory: Panopoulou 145, dept Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1235
            Fax: +49 30 8413 1128
            Email: panopoul@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones giving the same hybridisation pattern with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per cluster is selected for sequencing. The size of each
            cluster is an indicator of the frequency of a transcript in the
            analysed library. The cluster size as well the coordinates of the
            rest of the clones assigned to the same fingerprint cluster as the
            clone from which the above EST is generated is available at the
            amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
```

```
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="MPMGp531H0446"
                     /tissue_type="whole embryo"
                     /dev_stage="26 hrs (neurula stage)"
                     /clone_lib="Amphioxus 26hr cDNA library (Name convention:
                     BFL26 or MPMGp531)"
                     /note="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
                     EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT
                     primed and directionally cloned in pSport1 vector using a
                     NotI (5'-pGACTAGTTCTAGATGCGGCCGCCC(T)15-3' and a
                     SalI 5'- TCGACCCACGCGTCCG-3'adapters (Gibco BRL)."
ORIGIN
```

```
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (http://www.rzpd.de)
PCR Primers
FORWARD: 5' CCCCAGGGCTTACACTTTATCGTTCCGGGTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP)
Insert Length: 1  Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3', pSport3/86
High quality sequence stop: 661.
```

```
US-10-006-922A-12 (1-225) x BI385436 (1-661)

Alignment Scores:
Pred. No.:                 2.12e-16          Length:         661
Score:                     225.00            Matches:        62
Percent Similarity:        47.76%            Conservative:   34
Best Local Similarity:     30.85%            Mismatches:     89
Query Match:               18.53%            Indels:         16
DB:                        3                 Gaps:           6
```

```
Qy    16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGlu 35
Db    65 ATCCACCTTCCACGGCTTCCATCAACGGGCACGAGTTCGAACTGGTGGGCGAGGACAAGGC 124

Qy    36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
Db   125 GACCCGAACGGCCGGCCTCGCGTGACCACCAGGCGAAATCCACCAAG---GGTCCCTGAAG 181

Qy    56 PheAlaTrpAspAlaMetGlnPheGlyTyrGlySerLysSerValTyrValLysHis 75
Db   182 TTCTCTCCCTACTTGATGATCCCCCACCTCGGGTTACGGGTACTACCAGTACCTCCCTAC 241

Qy    76 ProAlaAspIleProAspTyrLysValSerPheProGlu-----GlyPheLysTrp 93
Db   242 CCG---GACGGACCCTCGCCTTTCCAGCCACCATGTTGGAAGGATTCAGGTATGCCAGTC 298

Qy    94 GluArgValMetAsnPheGluAspGlyValValThrValThrGlnAspSerSerLeu 113
Db   299 CATCGCGTGTTCGACTTTGAACGACGGAGGCCAAGCTGTCCATCGAGTTTAAATACTCCTAC 358

Qy   114 GlnAspGlyCysPheIleTyrLysValLysSerPheGlyValAsnPheProSerAspGly 133
Db   359 GAGGGGTTCCATATCAAGGCGGCGACATGAAGTTCACGGGAACCGGTTTCCCTGAGGACGGG 418

Qy   134 ProValMetGlnLeuLysThrMetGlyThrGlyTrpGluValAlaSer---ThrGluArgLeuTyrPro 152
Db   419 CCGGTCATGACCAGCCAGATTGCACCAGGACCGCTGCGTGTCCAAGAACACCTACTT 478

Qy   153 ArgAspGlyValLeuLysGlyGluIleHisLysLysAlaLeuLysLeuLysAspGlyHis 172
Db   479 AACGACAACACCATCGTGGACAGCGTTCAGCTGGACTTACAACCTGCAGAATGGGAAGCGC 538

Qy   173 TyrLeuValGluPheLysSerIleTyr----- 182
Db   539 TACAAGGGCCCAAGTCAGCGAGCCACTACATCTTCGGCAAGCCCTTCCCAGCGGACTTCATG 598

Qy   183 AlaLysLysProValGlnLeuLeuProGlyTyrTyrTyrValAsp---SerLysLeuAspIle 201
```

```
Db   599 AAGAAGCAGCCGGTCGCGCTTCGTGTACCGCAAGTGCTACGGTGAACGCTTCCAAGACCGAAATC 658
         |||||::||::  ||||        :::         ||::|        ||||::||::||:|||
Qy   202 Thr                                                              202
         |||
Db   659 ACC 661
```

RESULT 4

LOCUS	BW802012
DEFINITION	BW802012 Amphioxus Branchiostoma floridae mRNA unpublished cDNA library, larva whole animal Branchiostoma floridae cDNA clone bbiv023d19 5', mRNA sequence.
ACCESSION	BW802012
VERSION	BW802012.1 GI:66410228
KEYWORDS	EST.
SOURCE	Branchiostoma floridae
ORGANISM	Branchiostoma floridae
	Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
REFERENCE	1 (bases 1 to 508)
AUTHORS	Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
TITLE	Expressed genes in Branchiostoma floridae
JOURNAL	Unpublished (2005)
COMMENT	Contact: Tadasu Shin-i
	Center For Genetic Resource Information
	National Institute of Genetics
	1111 Yata, Mishima, Shizuoka 411-8540, Japan
	Tel: 81-559-81-6856
	Fax: 81-559-81-6855
	Email: tshini@genes.nig.ac.jp.
FEATURES	Location/Qualifiers
source	1..508
	/organism="Branchiostoma floridae"
	/mol_type="mRNA"
	/db_xref="taxon:7739"
	/clone="bbiv023d19"
	/tissue_type="whole animal"
	/dev_stage="larva"
	/clone_lib="Amphioxus Branchiostoma floridae cDNA library, larva whole animal"

ORIGIN

US-10-006-922A-12 (1-225) x BW802012 (1-508)

Alignment Scores:

Pred. No.:	8.90e-16	Length:	508
Score:	218.50	Matches:	52
Percent Similarity:	51.41%	Conservative:	21
Best Local Similarity:	36.62%	Mismatches:	50
Query Match:	18.00%	Indels:	19
DB:	5	Gaps:	4

```
Qy    18 MetGlyGlyThrValAlaAsnGlyHisGluPheGluIleGluGlyGlyGlyGlyArgPro 37
Db   103 ATGTTCGGCGTTCATCAACCGACCACGAGTTCGACCCTCGGTGGTGGTGGCGGACCCG 162
Qy    38 TyrGluGlyHisAsnThrValLysLeuValThrLeuLysSerGlyGlyProLeuProAla 57
Db   163 AATGACGGACGACCTGGAGACCAAAGTGCGGTACCTAAGGAGCC---CTGCCCTTCTCC 219
Qy    58 TrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAla 77
Db   220 CCCGTGATCCTGGCCCCTAACCTGGGGTACCAGTACCTGCCCCTTCCCGGCC 279
Qy    78 AspIleProAspTyrLysGlyLeuSerProPheProGlyPheTyrTrpGluArgValMet 97
Db   280 GGGACCTACACCGTACCAGCAGGCCATCACCACGGGGTACCAAAAGCACCGCACGGTTC 339
Qy    98 AsnPheGluAsnGlyLysGlyValValThrGlnAspSerSerLeuGlnAspGlyCys 117
Db   340 AAGTTCGAGGACGGCGGCGGGGTCATGACCATCAAC---------------------- 372
```

```
Qy   118 PheIleTyr--------------LysValLys----------PheGlyValAsn 128
Db   373 TTCCGCTACACCACTACTCAGGAGCAAGATCAAGGGAGGAGTTCCACGTGGTTGGACTCCGGG 432
Qy   129 PheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSerThrGlu 148
Db   433 TTCCCTGATGACGGGCCCTGTGATGACCAACTCACTCACCAGCAGCAGGATCACAACGTGGAG 492
Qy   149 ArgLeu 150
Db   493 AGGCTG 498
```

RESULT 5

LOCUS	BW736869
DEFINITION	BW736869 Amphioxus Branchiostoma floridae mRNA unpublished cDNA library, egg whole animal Branchiostoma floridae cDNA clone bbeg003b16 5', mRNA sequence.
ACCESSION	BW736869
VERSION	BW736869.1 GI:66323499
KEYWORDS	EST.
SOURCE	Branchiostoma floridae (Florida lancelet)
ORGANISM	Branchiostoma floridae
	Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
REFERENCE	1 (bases 1 to 646)
AUTHORS	Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
TITLE	Expressed genes in Branchiostoma floridae
JOURNAL	Unpublished (2005)
COMMENT	Contact: Tadasu Shin-i
	Center For Genetic Resource Information
	National Institute of Genetics
	1111 Yata, Mishima, Shizuoka 411-8540, Japan
	Tel: 81-559-81-6856
	Fax: 81-559-81-6855
	Email: tshini@genes.nig.ac.jp
	If you want to have a cDNA clone for this EST or if you have any
	questions, please send an e-mail to Nori Satoh
	(satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka Satou
	(yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES	Location/Qualifiers
source	1..646
	/organism="Branchiostoma floridae"
	/mol_type="mRNA"
	/db_xref="taxon:7739"
	/clone="bbeg003b16"
	/tissue_type="whole animal"
	/dev_stage="egg"
	/clone_lib="Amphioxus Branchiostoma floridae cDNA library, egg whole animal"

ORIGIN

US-10-006-922A-12 (1-225) x BW736869 (1-646)

Alignment Scores:

Pred. No.:	1.26e-15	Length:	646
Score:	218.50	Matches:	54
Percent Similarity:	48.28%	Conservative:	30
Best Local Similarity:	31.03%	Mismatches:	85
Query Match:	18.00%	Indels:	5
DB:	5	Gaps:	4

```
Qy    16 ValArgMetGluGlyThrValAlaAsnGlyHisGluPheGluIleGluGlyGlyGlyGly 35
Db    61 ATCGACCATTCACGGCTCCATCAACGGCCCACGAGTTCGACATGGTGGGGGGAGGAAAAGGC 120
Qy    36 ArgProTyrGluGlyHisAsnThrValLysLeuValThrLeuArgSerGlyGlyProLeuPro 55
Db   121 GACCCGAACCGCGGTCGGTCGGTGGTGACCAGCGGAAATCCACCAAG---GGTCCCGTAGG 177
Qy    56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
Db   178 TTCTCTCCCACTTGATGATCCCCACTCGGGTACGGGTACTACCACCAGTACCTCCCTAC 237
```

RESULT 6

```
Qy   76 ProAlaAspIleProAspTyrLysLysLeuSerPheProGlu------GlyPheLysTrp  93
           :::        |||||||||       |  |       ||
Db  238 CCG---GACCGACCCTCGGCCTTTCCAGGCCACCATGTTGGAAGGATCGGGGTATACAGTC 294

Qy   94 GluArgValMetAsnPheGluAspGlyIleGlyValValThrValThrGlnAspSerLeu 113
Db  295 CACCGGCGTGTTTGACTTCGAAGACGGAGCCAAGCTGTCCATCGAGTTCAAGTACTCCTAC 354

Qy  114 GlnAspGlyCysPheIleTyrLysValValAlaAsnPheProSerAspGly 133
Db  355 GAGGGGTTCCCATATTAAGGCCCGAAGTTCACGGAGGACACCGGTTCCCTGAGGACGGG 414

Qy  134 ProValMetGlnLysGlyThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro 152
Db  415 CCGGTCATGACCAGCCAGATTGTCGCGTGTCCAAGAACACCTACCTG 474

Qy  153 ArgAspGlyValLeuLysGlyGluIleLeuHisLysAlaLeuLysLeuLysAspGlyHis 172
Db  475 AACGACAACACCATCGGTGGACAACTTCGACTGGAACCTGCAGAATGGGAAGCGC 534

Qy  173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
Db  535 TACAGGGCCCGAGTGTCGAGCCACTACATCTTCGACAAGCCC 576
```

Alignment Scores:

Score:	1.64e-15
Pred. No.:	217.50
Percent Similarity:	48.85%
Best Local Similarity:	31.03%
Query Match:	17.92%
DB:	5

Length:	639
Matches:	54
Conservative:	31
Mismatches:	84
Indels:	5
Gaps:	4

RESULT 7

US-10-006-922A-12 (1-225) x BW736108 (1-639)

```
Qy   16 ValArgMetGluGlyTyrThrValAsnGlyHisGlyPheGlyIleGluGlyGluGlyGly  35
Db   67 ATCCACCTCACGGTCCCATCAAACGGCCACCAGTTCGACCTGGTGGGCGGGAGGAAAAGGC 126

Qy   36 ArgProTyrGluGlyHisAsnThrValLysLeuValThrLysGlyGlyProLeuPro  55
Db  127 GACCCGGAGCGCCGGCCGCTCGCGTGGTGACCACACGGAAATCCACCAGG---GGTCCCTGAAG 183

Qy   56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValThrValLysHis  75
Db  184 TTCTCTCCCTACTTGATGATCCCCACCCTCGGGGTGTACGGGTACTACCAGTACTCCCTAC 243

Qy   76 ProAlaAspIleProAspTyrLysLysLeuSerPheProGlu------GlyPheLysTrp  93
Db  244 CCG---GACCGACCCTCGGCCTTTCCAGGCCACCATGTTGGAAGGTATGCAGTC 300

Qy   94 GluArgValMetAsnPheGluAspGlyIleGlyValValThrValThrGlnAspSerLeu 113
Db  301 CACCGGCGTGTTCGACTTTGAAGACGGAGCCAAGCTGTCCATCGAGTTTAAGTACTCCTAC 360

Qy  114 GlnAspGlyCysPheIleTyrLysValValAlaAsnPheProSerAspGly 133
Db  361 GAGGGGTTCCCATATCAAGGCCGGACAGAGTTCACGGAGGACACCGGTTCCCTGAGGACGGG 420

Qy  134 ProValMetGlnLysGlyThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro 152
Db  421 CCGGTCATGACCAGCCAGATTATCGACACGGCGTGCGTGTCCAAGAACAACCTACCTT 480

Qy  153 ArgAspGlyValLeuLysGlyGluIleLeuHisLysAlaLeuLysLeuLysAspGlyHis 172
Db  481 AACGACAACACCATCGGTGGACAGCTTCGACCTGGAACCTGCAGAATGGGAAGCGC 540

Qy  173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
Db  541 TACAAGGGCCCGAGTGACGAGCCACTACATCTTCGGCAAGCCC 582
```

Left column

```
ORIGIN
        /dev_stage="egg"
        /clone_lib="Amphioxus Branchiostoma floridae
        cDNA library, egg whole animal"

US-10-006-922A-12 (1-225) x BW738638 (1-604)

Alignment Scores:
  Pred. No.:                 2.0le-15
  Score:                     216.50        Length:        604
  Percent Similarity:        48.85%        Matches:        54
  Best Local Similarity:     31.03%        Conservative:   31
  Query Match:               17.83%        Mismatches:     84
  DB:                        5             Indels:          5
                                           Gaps:            4
```

```
Qy   16 ValArgMetGluGlyTyrThrValAsnGlyHisGluPheGluIleGluGlyGlyGluGlyGly  35
         :::       ||||||     ||||||||||||||:||||||||
Db   67 ATCCACCTTCACGGCTCCATCAACGGCCACAGAGTTCGACATGGTGGGCGGAGGAAAAGGC 126

Qy   36 ArgProTyrGluGlyHisAsnThrValLysLeuLysGlyValThrGlyGlyProLeuPro   55
         |||   |||||     ||  ::||||||||  ||  ::: ||
Db  127 GACCCGAACGCCGGCTCGTGGTGACCACAGCGAAATCCACCAAG---GGTCCCCTGAAG  183

Qy   56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLeuLysValTyrValHis   75
         ||| :::        |||     ||||     :::  ::: |||
Db  184 TTCTCCTCCTTACTTGATGATCCCCACCTCGGGTTACGGCGTACTACCAGTACCTCCCTAC 243

Qy   76 ProAlaAspIleProAspTyrLysGlySerPheProGlu------GlyPheLysTrp     93
         |||       :::        |||:::    ||         ::: |||
Db  244 CCG---GACGGACCCTCGCCTTTCAGGCGATTCCAGGGT            GGGFheLysTrp 300

Qy   94 GluArgValMetAsnPheGluGlyHisAspGlyValValThrGlnAspSerSerLeu    113
         |||   :::        ||||||     :::  ::: |||
Db  301 CATCGCGTGTTCGACTTTGAAGACGGAGGGGGTGTCCATCGAGTTTAGTACTTCCTAC 360

Qy  114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAlaAsnPheProSerGly  133
         ||| :::        ||||||     :::  ::: |||
Db  361 GAGGGTTCCCATATCAAGGCCGACATGAAGTTCAGGGGTTTCCCTGAGGACGCGG  420

Qy  134 ProValMetGlnLysLysThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro  152
         |||   :::        ||||||     :::  ::: |||
Db  421 CCGGTCATGACAAGCCAGATTGTCGACCAGGACGCCGTGTCCAAGAACACCTACCTC  480

Qy  153 ArgAspGlyValLeuLysGlyLeuIleHisGlyAlaLeuLysLeuLysAspGlyGlyHis 172
         |||       :::        |||:::    ||
Db  481 AACGACAACACCATCGTGACAGCTTCGACTGGACTTACAATCTGCAGAACGGAAGCGC 540

Qy  173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro                   186
         |||       :::        |||
Db  541 TACAGGCCAAAGTGACGAGCCACTACATCTTCGCAAGCC                     582
```

```
RESULT 8
LOCUS       BW741633      606 bp   mRNA    linear   EST 09-AUG-2005
DEFINITION  BW741633 Amphioxus Branchiostoma floridae unpublished cDNA library,
            egg whole animal Branchiostoma floridae cDNA clone bbeg040g05 5',
            mRNA sequence.
ACCESSION   BW741633
VERSION     BW741633.1  GI:66328281
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
ORGANISM    Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
```

REFERENCE   1 (bases 1 to 606)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE   Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)

Right column

```
FEATURES             Location/Qualifiers
     source          1..606
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /clone="bbeg040g05"
                     /db_xref="taxon:7739"
                     /clone_lib="Amphioxus Branchiostoma floridae
                     cDNA library, egg whole animal"
                     /dev_stage="egg"
                     /tissue_type="whole animal"

ORIGIN

US-10-006-922A-12 (1-225) x BW741633 (1-606)

Alignment Scores:
  Pred. No.:                 2.0le-15
  Score:                     216.50        Length:        606
  Percent Similarity:        48.28%        Matches:        54
  Best Local Similarity:     31.03%        Conservative:   30
  Query Match:               17.83%        Mismatches:     85
  DB:                        5             Indels:          5
                                           Gaps:            4
```

```
Qy   16 ValArgMetGluGlyTyrThrValAsnGlyHisGluPheGluIleGluGlyGlyGluGlyGly  35
         :::       ||||||     ||||||||||||||:||||||
Db   69 ATCCACCTTCACGGCTCCATCAACGGCCACAGAGTTCGACATGGTGGGCGGGGGAGGAAAGGC 128

Qy   36 ArgProTyrGluGlyHisAsnThrValLysLeuLysGlyValThrGlyGlyProLeuPro   55
         |||   |||||     ||  ::||||||||  ||  ::: ||
Db  129 GACCCGAACGCCGGCTCGTGGTGACCACAGCGAAATCCACCAAG---GGTCCCCTGAAG  185

Qy   56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLeuLysValTyrValHis   75
         ||| :::        |||     ||||     :::  ::: |||
Db  186 TTCTCCTCCTTACTTGATGATCCCCACCTCGGGTTACGGCGTACTACCAGTACCTCCTAC 245

Qy   76 ProAlaAspIleProAspTyrLysGlySerPheProGlu------GlyPheLysTrp     93
         |||       :::        |||:::    ||         ::: |||
Db  246 CCG---GACGGACCCTCGCCTTTCAGGCGATTCCAGGGT            GGGPheLysTrp 302

Qy   94 GluArgValMetAsnPheGluGlyHisAspGlyValValThrGlnAspSerSerLeu    113
         |||   :::        ||||||     :::  ::: |||
Db  303 CACCGCGTGTTCGATTTTGAAGGACGGAGGGGGTGTCCATCGAGTTTAAGTACTTCCTAC 362

Qy  114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAlaAsnPheProSerGly  133
         ||| :::        ||||||     :::  ::: |||
Db  363 GAGGGTTCCCATATTAAGGCCGACATGAAGTTCACGGGACCGGTTTCCCTGAGGACGGC  422

Qy  134 ProValMetGlnLysLysThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro  152
         |||   :::        ||||||     :::  ::: |||
Db  423 CCGGTCATGACCAGCCAGATTGTGCGACCAGGACGGCGTGTCCAAAGAACACCTACCTC  482

Qy  153 ArgAspGlyValLeuLysGlyLeuIleHisGlyAlaLeuLysLeuLysAspGlyGlyHis 172
         |||       :::        |||:::    ||
Db  483 AACGACAACACCATCGTGACAGCTTACGACACTTCGACTTACAATCTGCAGAACGGAAAGCGC 542

Qy  173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro                   186
         |||       :::        |||
Db  543 TACAGGACCCGAGTGACGAGCCACTACATCTTCGACAAGCC                    584
```

```
RESULT 9
LOCUS       BI379468      637 bp   mRNA    linear   EST 26-AUG-2003
DEFINITION  BI379468 BPLG1_001099 Amphioxus 5-6 hrs cDNA library (Name
            convention: BPLG or MPMGp498) Branchiostoma floridae cDNA clone MPMGp49801918 5',
            mRNA sequence.
ACCESSION   BI379468
VERSION     BI379468.1  GI:30914617
KEYWORDS    EST.
```

SOURCE   Branchiostoma floridae (Florida lancelet)
  ORGANISM Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1 (bases 1 to 637)
  AUTHORS   Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
            Herwig,R., Vingron,M. and Lehrach,H.
  TITLE   New evidence for genome-wide duplications at the origin of
            vertebrates using an amphioxus gene set and completed animal
            genomes
  JOURNAL   Genome Res. 13 (6A), 1056-1066 (2003)
   PUBMED   12799346
  COMMENT   Contact: Panopoulou G
            laboratory 145, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1235
            Fax: +49 30 8413 1128
            Email: panopoul@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones giving the same hybridisation pattern with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per cluster is selected for sequencing. The size of each
            cluster is an indicator of the frequency of a transcript in the
            analysed library. The cluster size as well the coordinates of the
            rest of the clones assigned to the same fingerprint cluster as the
            clone from which the above EST is generated is available at the
            amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
            Clones and filters are distributed via the Resource Center/Primary
            Database of the German Genome Project (http://www.rzpd.de)
            PCR PRimers
            FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP)
            BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP)
            Insert Length: 1200   Std Error: 0.00
            Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3' pSport3/86
            High quality sequence stop: 637.
FEATURES   Location/Qualifiers
     source   1..637
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="MPMGp49801918"
                     /tissue_type="whole embryo"
                     /dev_stage="5-6 hrs (gastrula stage)"
                     /lab_host="E.coli, XL1 blue"
                     /clone_lib="Amphioxus 5-6 hrs cDNA library (Name
                     convention: BFLG or MPMGp498)"
                     /note="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
                     Site 2: NotI, BamHI, HindIII (3'); Oligodt primed and
                     directionally cloned in pSport1 vector using a NotI
                     (5'-pGACTAGTTCTAGATCGGCGAACGGCGCGCC (T)15-3' and a SalI 5'-
                     TCGACCACGCGTCCG-3'adapters (Gibco BRL)."

ORIGIN

Alignment Scores:
Pred. No.:
Score: 2.16e-15
Percent Similarity: 216.50
Best Local Similarity: 48.57%
Query Match: 30.29%
DB: 17.83%
                    3

US-10-006-922A-12 (1-225) x BI379468 (1-637)

Qy   15 LysValArgMetGlyGlyThrValAsnGlyHisGluPheGluPheGluGlyGluGluGlu 34
           ||||| ::: ||||||||||||:: ||||||||||| ::: :::
Db   79 GAGATCCAACTTCCACGGTTCCATCACACGGGCACGAGTTCGACCTGGCTGGTGTGGAAAA 138

Qy   35 GlyArgProTyrGluLyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeu 54
           |||| :::||||||| ||||| ::: ||||||| :::   :::
Db   139 GGGCGACCCAAAACGCGGGCTTCGCTGGTGTGGTGACCACAGCAGAAATCACCCAG---GGTCCCCTG 195

RESULT 10
BI387917
LOCUS   BI387917   687 bp   mRNA   linear   EST 26-AUG-2003
DEFINITION BFL26_002620 Amphioxus 26hr cDNA library (Name convention: BFL26 or
            MPMGp531) Branchiostoma floridae cDNA clone MPMGp531N0767 5', mRNA
            sequence.
ACCESSION   BI387917
VERSION   BI387917.1 GI:30922753
KEYWORDS   EST.
SOURCE   Branchiostoma floridae (Florida lancelet)
  ORGANISM Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1 (bases 1 to 687)
  AUTHORS   Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
            Herwig,R., Vingron,M. and Lehrach,H.
  TITLE   New evidence for genome-wide duplications at the origin of
            vertebrates using an amphioxus gene set and completed animal
            genomes
  JOURNAL   Genome Res. 13 (6A), 1056-1066 (2003)
   PUBMED   12799346
  COMMENT   Contact: Panopoulou G
            laboratory 145, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1235
            Fax: +49 30 8413 1128
            Email: panopoul@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones giving the same hybridisation pattern with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per cluster is selected for sequencing. The size of each
            cluster is an indicator of the frequency of a transcript in the
            analysed library. The cluster size as well the coordinates of the
            rest of the clones assigned to the same fingerprint cluster as the
            clone from which the above EST is generated is available at the
            amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
            Clones and filters are distributed via the Resource Center/Primary
            Database of the German Genome Project (http://www.rzpd.de)
            PCR PRimers
            FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP)
            BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP)
            Insert Length: 1   Std Error: 0.00
            Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3' pSport3/86

Qy   55 ProPheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLys 74
           ||| |:::   :::
Db   196 AAGTTCTCTCCCCACTTGATGATCCCCACCTCGGGTACGGGTACTACCAGTACCTCCCC 255

Qy   75 HisProAlaAspIleProAspTyrLysLysLeuSerPhePheGlu------GlyPheLys 92
           :::|||||||||||||| :::   :::   :::
Db   256 TACCCG---GACGGACCTCGCCCTTCCAGGCCACCCATGTTGGAGGATCGGGGTATACA 312

Qy   93 TrpGluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAsnSerSer 112
           :::||||||||||||||||| ||| ::: :::
Db   313 GTCCACCGGTCGTTTGACTTCGAAGACGGAGGCAAGCTGTCCATCGAGTTCAAGTACTCC 372

Qy   113 LeuGlnAspGlyCysPheIleTyrLysValHisPheIleGlyValAlaAsnPheProSerAsp 132
           :::
Db   373 TACGAGGGTTCCCATATTAAGCCCGACATGAAGTTCACGGAGAACCGGTTTCCCTGAGGAT 432

Qy   133 GlyProValMetGlnLysLysThrMetGlyTrpGluValSer---ThrGluArgLeuTyr 151
           ||||||| ::: ||| ::: :::
Db   433 GGGCCGGTCATGACCAGCCAGATTGTCGACCAGGAGCGGCTGGGTGTCCAAGAACACCTAC 492

Qy   152 ProArgAspGlyValLeuLysGlyGlyIleHisLysAlaLeuLeuLysAspGlyGly 171
           ::: |||||||||||||||||||
Db   493 CTCAACGACACAACCATCGTGGACACTTCGACTGGACTTACAACCTGCAGAATGGGAAG 552

Qy   172 HisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
           |||
Db   553 CGCTACAGGGCCGAGTGACCAGGACCACTACACATCTTCGACACAGCCC 597

First record

```
FEATURES             Location/Qualifiers
source               High quality sequence stop: 687.
                     1..687
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="MPMGp531N07767"
                     /tissue_type="whole embryo"
                     /dev_stage="26 hrs (neurula stage)"
                     /lab_host="E.coli, XL1 blue"
                     /clone_lib="Amphioxus 26hr cDNA library (Name convention:
                     BFL26 or MPMGp531)"
                     /note="Vector: pSport1 (Gibco BRL); Site_1: SalI, KpnI,
                     EcoRI (5'); Site_2: NotI, BamHI, HindIII (3'); OligodT
                     primed and directionally cloned in pSport1 vector using a
                     NotI (5'-pGACTAGTTCTAGATCGGCCGGCCGCGCCC(T)15-3' and a
                     SalI 5'- TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN
```

US-10-006-922A-12 (1-225) x BI387917 (1-687)

Alignment Scores:

Pred. No.:	2.41e-15	Length:	687
Score:	216.50	Matches:	54
Percent Similarity:	48.85%	Conservative:	31
Percent Identity:	31.03%	Mismatches:	84
Best Local Similarity:	17.83%	Indels:	5
Query Match:	3	Gaps:	4
DB:			

```
Qy   16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGluGly 35
         |||:||||||||:||||:|||||||||||||:||||||:|:::||
Db   99 ATCCAACCTCACGGCTCCATCAACGGCCACGAGTTCGACATGGTGGGGGAGGAAAGGC 158

Qy   36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
         |||||||||||:||:|||||||||:|:::||
Db  159 GACCCGAACGCCGGCTCGCTGGTGACCACAGCGAAATCCACCAAG---GGTCCCCTGAAG 215

Qy   56 PheAlaTrpAspIleLeuSerProGlnPheGlnIleTyrGlySerLysValTyrValLysHis 75
         :: |||                     |||
Db  216 TTCTCCTCCTACTTGATGATCCCACCTCGGGTACGGCTACTACCAGTACTCCCCTAC 275

Qy   76 ProAlaAspIleProAspTyrLysGlyLysSerLeuSerPheProGlu------GlyPheLysTrp 93
         :::                   |||      |||||
Db  276 CCG---GACGGAGCCCTCGCCCTTCCAGGCCCACATCCGGTTCCCTGAGGACCAGTC 332

Qy   94 GluArgValMetAsnPheGluLeuAspGlyValValThrValThrGlnAspSerSerLeu 113
         ||||||                   ::||:
Db  333 CATCGCGTCGTTCGACTTTGAAGACCGAGGCAAGCTGTCCATCGAGTTTGAGTACTCCTAC 392

Qy  114 GlnAsnGlyCysPheIleTyrIleGlyValValSerPheIleGlyValAlaAsnPheArgSerAspGly 133
         :::                   ||||||   :::
Db  393 GAGGGTTCCCATATCAAGGGCACATGAAGTTCACGGGAACCGGTTTCCCTGAGGACGGG 452

Qy  134 ProValMetGlnIleLysLysThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro 152
         :::                   |||      :::
Db  453 CCGGTCATGACCAGGCAGAATTGTCGACCAAGGACACCTACCTT 512

Qy  153 ArgAspGlyValLeuLysGlyIleHisLysAlaLeuLysAspGlyGlyHis 172
         :::                   |||      :::
Db  513 AACGACAACCACCATCGTGACCAGCTTCGACTGGACTTACAACCTGCAGAATGGGAAGCC 572

Qy  173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
         ||||||                   :::
Db  573 TACAAGGGCCCAAGTGACGAGCCACTACACTCTTCGCCAAGCCC 614
```

Second record

```
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
ORGANISM    Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
            Herwig,R., Vingron,M. and Lehrach,H.
TITLE       New evidence for genome-wide duplications at the origin of
            vertebrates using an amphioxus gene set and completed animal
            genomes
JOURNAL     Genome Res. 13 (6A), 1056-1066 (2003)
PUBMED      12799346
COMMENT     Contact: Panopoulou G
            Laboratory 145, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1235
            Fax: +49 30 8413 1128
            Email: panopoul@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones giving the same hybridisation pattern with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per cluster is selected for sequencing. The size of each
            cluster is an indicator of the frequency of a transcript in the
            analysed library. The cluster size as well the coordinates of the
            rest of the clones assigned to the same fingerprint cluster as the
            clone from which the above EST is generated is available at the
            amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
            Clones and filters are distributed via the Resource Center/Primary
            Database of the German Genome Project (http://www.rzpd.de)
            PCR PRimers
            FORWARD: 5' CCCCAGGCGTTACACTTTATGCTTCCGGCTCG 3' (M13RSP)
            BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP)
            Insert Length: 1   Std Error: 0.00
            Seq Primer: 5'-CCGGTCCGGAATTCCCGGGT-3' pSport3/86
            High quality sequence stop: 643.

FEATURES             Location/Qualifiers
source               1..643
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="MPMGp531L1370"
                     /tissue_type="whole embryo"
                     /dev_stage="26 hrs (neurula stage)"
                     /lab_host="E.coli, XL1 blue"
                     /clone_lib="Amphioxus 26hr cDNA library (Name convention:
                     BFL26 or MPMGp531)"
                     /note="Vector: pSport1 (Gibco BRL); Site_1: SalI, KpnI,
                     EcoRI (5'); Site_2: NotI, BamHI, HindIII (3'); OligodT
                     primed and directionally cloned in pSport1 vector using a
                     NotI (5'-pGACTAGTTCTAGATCGGCCGGCCGCGCCC(T)15-3' and a
                     SalI 5'- TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN
```

US-10-006-922A-12 (1-225) x BI386699 (1-643)

Alignment Scores:

Pred. No.:	6.69e-15	Length:	643
Score:	212.50	Matches:	56
Percent Similarity:	47.87%	Conservative:	34
Percent Identity:	29.79%	Mismatches:	87
Best Local Similarity:	17.50%	Indels:	11
Query Match:	3	Gaps:	5
DB:			

```
Qy   16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGluGly 35
         :|:
Db   81 ATCCAACCTCACGGCTCCATCAACGGCCACGAGTTCGACATGGTGGGGGAGGAAAGGC 140

Qy   36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
         |||||||||||:||:|||||||||:|:::||
Db  141 GACCCGAACGCCGGCTCGCTGGTGACCACAGCGAAATCCACCAAG---GGTCCCCTGAAG 197
```

Left-margin entry

```
RESULT 11
LOCUS       BI386699    643 bp    mRNA    linear    EST 26-AUG-2003
DEFINITION  BFL26_001402 Amphioxus 26hr cDNA library (Name convention: BFL26 or
            MPMGp531) Branchiostoma floridae cDNA clone MPMGp531L1370 5', mRNA
            sequence.
ACCESSION   BI386699
VERSION     BI386699.1  GI:30921604
```

RESULT 12

```
Qy    56  PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
              ||||||        |||||        :::
Db   198  TTCTCCCCTACTTGATGATCCCCCCAACCTCGGGTACGGGTACCAGTACCTCCCCTAC 257

Qy    76  ProAlaAspIleProAlaSpTyrLysGlyLeuSerPheProGlu------GlyPheLysTrp 93
              :::             ||||        :::
Db   258  CCG---GACGGACCCTCGCCTTTCAGGCCTTCCAGTGTTGGAAGGAATCGGGGGTATGCAGTC 314

Qy    94  GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
              ||||||        |||||||        |||
Db   315  TACCGGCGTGTTCGACTTTGAAGACGGAGGCAAGCTACCAGTTTAAGTACTATCCTAC 374

Qy   114  GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAlaAsnPheProSerSerLeu 133
              ||||||        |||||        :::
Db   375  GAGGGTTCCCATATCAAGGCCGACATGAAGCTGATGGGACACGGGTTTCCCTGACGACGGC 434

Qy   134  ProValMetGlnLysLysThrMetGlyTrpTrpGluAlaSer---ThrGluArgLeuTyrPro 152
              ||||||        :::        :::
Db   435  CCAGTCATGACCAGCCAGAGATTGTGCACAGCGACGCTCGCCGTGTTCCAAGAGACGGTATCTT 494

Qy   153  ArgAspGlyValLeuLysGlyLeuIleHisLysAlaLeuLysLysAspGlyHis 172
              |||        :::
Db   495  AACACACAACCACCGTGGACAGCTTGGAGTTACAACCTGACGAATGGGAAGCGC 554

Qy   173  TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro----------- 186
              |||        :::
Db   555  TACAGGCGCCGAGTGTTCNAGCCACTACATCTTCGACAAGCCCTTTTCAGCCGCCGGATCTCATG 614

Qy   187  ValGlnLeuProGlyTyrTyrTyr 194
              |||
Db   615  ATAAAGCAGACGCGGTTCTTCGTGGTAC 638
```

RESULT 13

```
US-10-006-922A-12 (1-225) x BI382638 (1-675)

Qy    16  ValArgMetGluGlyThrValAsnGlyValHisGluPheGluIleGluGlyGluGlyGly 35
              ||||||        |||||||        ||||||
Db   110  ATCCACCTTCACGGCTCCATCAACGGCCCACGAGTTCGACATGGTGGGGGGGGAGGAAGGC 169

Qy    36  ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
              |||        |||
Db   170  GACCCGAACGCCGGCTTCCGTGTGACCAAGCAGAAATCCACCAAG---GGTCCCCTGAAG 226

Qy    56  PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
              ||||||        :::        :::
Db   227  TTCTCCCCTACTTGATGATCCCCCCAACCTCGGGTACGGGTACCAGTACCTCCCCTAC 286

Qy    76  ProAlaAspIleproAspTyrLysGlyLeuSerPheProGlu------GlyPheLysTrp 93
              :::        :::
Db   287  CCG---GACGGACCCTCGCCTTTCAGGTCTTCCAGTGTTGGAAGGAATCGGGGGTATGCAGTC 343

Qy    94  GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
              ||||||
Db   344  TACCGGCGTGTTCGACTTTGAAGACGGAGGCAAGCTGACTACCGAGTTTAAGTACTACTCCTAC 403

Qy   114  GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAlaAsnPheProSerSerLeu 133
              |||
Db   404  GAGGGTTCCCATATCAAGGCCGACATGAAGCTGATGGGAGACGGGTTTCCCTGACGACGGC 463

Qy   134  ProValMetGlnLysLysThrMetGlyTrpTrpGluAlaSer---ThrGluArgLeuTyrPro 152
              ||||||        :::
Db   464  CCAGTCATGACCAGCCAGAGATTGTGCACAGCGACGCTCGCCGTGTTCCAAGAGACGGTATCTT 523

Qy   153  ArgAspGlyValLeuLysGlyLeuIleHisLysAlaLeuLysLysAspGlyHis 172
              |||
Db   524  AACACACAACCACCGTGGACAGCTTGGAGTTACAACCTGACGAATGGGAAGCGC 583

Qy   173  TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
              |||
Db   584  TACAGGCGCCGAGTGTTCNAGCCACTACATCTTCGACAAGCCC 625
```

LOCUS   BW883955   590 bp   mRNA   linear   EST 24-MAY-2005

LOCUS   BI382638   675 bp   mRNA   linear   EST 26-AUG-2003
DEFINITION BFLG2_000815 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
            or MPMGp498) Branchiostoma floridae cDNA clone MPMGp498F0258 5',
            mRNA sequence.
ACCESSION   BI382638
VERSION   BI382638.1   GI:30917704
KEYWORDS   EST.
SOURCE   Branchiostoma floridae (Florida lancelet)
  ORGANISM Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1 (bases 1 to 675)
  AUTHORS   Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
            Herwig,R., Vingron,M. and Lehrach,H.
  TITLE   New evidence for genome-wide duplications at the origin of
            vertebrates using an amphioxus gene set and completed animal
            genomes
  JOURNAL   Genome Res. 13 (6A), 1056-1066 (2003)
  PUBMED   12799346

COMMENT   Contact: Panopoulou G
            Max-Planck-Institut fuer Molekulare Genetik
            laboratory 145, dept.Lehrach
            Ihnestr 63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1235
            Fax: +49 30 8413 1128
            Email: panopoulo@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones giving the same hybridisation pattern with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per cluster is selected for sequencing. The size of each
            cluster is an indicator of the frequency of a transcript in the
            analysed library. The cluster size as well the coordinates of the
            clone from which the above EST is generated is available at the
            rest of the clones assigned to the same fingerprint cluster as the
            amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
            Clones and filters are distributed via the Resource Center/Primary
            Database of the German Genome Project (http://www.rzpd.de)

FEATURES   Location/Qualifiers
     source   1..675
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="MPMGp498F0258"
                     /tissue_type="whole embryo"
                     /dev_stage="5-6 hrs (gastrula stage)"
                     /lab_host="E.coli, XL1 blue"
                     /clone_lib="Amphioxus 5-6 hrs cDNA library (Name
                     convention: BFLG or MPMGp498)"
                     /note="Vector: pSport1; Site_1: SalI, KpnI, EcoRI
                     Site_2: NotI, BamHI, HindIII(3'); Oligo dT primed and
                     directionally cloned in pSport1 vector using a NotI
                     (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCC (T)15-3' and a SalI 5'-
                     TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Alignment Scores:
Pred. No.:
Score: 7.17e-15
Percent Similarity: 212.50   Length: 675
Best Local Similarity: 49.43%   Matches: 54
Query Similarity: 31.03%   Conservative: 32
Query Match: 17.50%   Mismatches: 83
DB: 3   Indels: 5
                                        Gaps: 4

US-10-006-922A-12 (1-225) x BI382638 (1-675)

PCR PRImers
FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCGCAGCTGGGCGAAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATTCCGGGGT-3' pSport3/86
High quality sequence stop: 675.
```

DEFINITION   BW883955 Amphioxus Branchiostoma floridae unpublished cDNA library, 5', mRNA sequence.
ACCESSION   BW883955
VERSION   BW883955.1 GI:66498632
KEYWORDS   EST.
SOURCE   Branchiostoma floridae (Florida lancelet)
ORGANISM   Branchiostoma floridae
             Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
             Branchiostoma.
REFERENCE   1 (bases 1 to 590)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE   Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6855
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES
  source   1..590
             /organism="Branchiostoma floridae"
             /mol_type="mRNA"
             /db_xref="taxon:7739"
             /clone="bbnel135j09"
             /tissue_type="whole animal"
             /dev_stage="neurula"
             /clone_lib="Amphioxus Branchiostoma floridae unpublished cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 210.50   Length: 590
Percent Similarity: 48.85%   Matches: 54
Best Local Similarity: 31.03%   Conservative: 31
Query Match: 17.34%   Mismatches: 84
                             1.04e-14   Indels: 5
                                               Gaps: 4

US-10-006-922A-12 (1-225) x BW883955 (1-590)

Qy   16 ValArgMetGluGlyThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGly 35
           || ::: ||||||||
Db   70 ATCACCTTCACGCGCTCCATCAACGACGGCCACGAGTTCGACAGTTGGGGGGAGGAAAGGC 129

Qy   36 ArgProTyrGluGlyHisAsnThrValLeuLysLeuValGlyThrLeuGlyProLeuPro 55
        ||||| |||||||||||| |||||||||||
Db 130 GACCCGAACGCGCGGCTCGGTGGTGACCACGAGAAATCCACCAAG---GGTCCCCTGAAG 186

Qy   56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
        |||||||| :: ||
Db 187 TTCTCTCCCTACTTGATGATCCCCAGACCCTGGGTACTACCAGTAC 246

Qy   76 ProAlaAspIleProAspTyrLysLysPheGlyGlyProLeu------GlyPheLysTrp 93
        |||
Db 247 CCG---GACGGAGACCCTCCATGGTTGAAGATCGGGAAGGATCGGGTATGCAGTC 303

Qy   94 GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
             |||||| :::|||||| ||| ||||||||
Db 304 TACCGCGTGTTCGACTTTGAAGACGGAGGGGTCGTCACCGTTAAGTACTCCTAC 363

Qy 114 GlnAspGlyCysPheIleTyrLysGlyValLysPheIleGlyValAsnPheProSerAsp 133
        |||
Db 364 GAGGGTACCCCATATCAAGGACGCATGAAGCTGATGGGAAGCGGTTTCCCTGAGGACGGC 423

Qy 134 ProValMetGlnLysLysThrMetGlyTrpGlnAlaSer---ThrGluArgLeuTyrPro 152
        ||| ||| | :: :::
Db 424 CCAGTCATGCAGAAGAAGACCAGGACGGCCTGGCTGGTCCAAGAAGAGCGTATCTT 483

Qy 153 ArgAspGlyValLeuLeuGlyGluIleHisLysAlaLeuLysLeuLysAspGlyHis 172
        ::: | ::: ||||||||||
Db 484 AACAACAACAACACCATCGTGGACAGCTTCGACGTGGAGTTACAACCTGCAGAATGGGGAAGCGGC 543

Qy 173 TyrLeuValGluPheLysSerIleTyrMetAlaLysGlyLysPro 186
        ||| || :::
Db 544 TACAGGGCCCGAGTGTGCGACGGCCACTACACATCTTCGACAAGCCC 585

RESULT 14
BW893733
LOCUS   BW893733   812 bp   mRNA   linear   EST 24-MAY-2005
DEFINITION   BW893733 Amphioxus Branchiostoma floridae unpublished cDNA library,
             5', mRNA sequence.
ACCESSION   BW893733
VERSION   BW893733.1 GI:66511577
KEYWORDS   EST.
SOURCE   Branchiostoma floridae (Florida lancelet)
ORGANISM   Branchiostoma floridae
             Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
             Branchiostoma.
REFERENCE   1 (bases 1 to 812)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE   Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6855
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES
  source   1..812
             /organism="Branchiostoma floridae"
             /mol_type="mRNA"
             /db_xref="taxon:7739"
             /clone="bbnel123k07"
             /tissue_type="whole animal"
             /dev_stage="neurula"
             /clone_lib="Amphioxus Branchiostoma floridae unpublished cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 210.50   Length: 812
Percent Similarity: 46.92%   Matches: 60
Best Local Similarity: 28.44%   Conservative: 39
Query Match: 17.34%   Mismatches: 91
                             1.63e-14   Indels: 21
                                               Gaps: 7

US-10-006-922A-12 (1-225) x BW893733 (1-812)

Qy   16 ValArgMetGluGlyThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGly 35
           || ::: |||||||| :::
Db   60 ATCACCTTCACGCGCTCCATCAACGACGGCCACGAGTTCGACAGTTGGGGGGAGGAAAGGC 119

Qy   36 ArgProTyrGluGlyHisAsnThrValLeuLysLeuValGlyThrLeuGlyProLeuPro 55
        ||||| |||||||||||| |||||||||||
Db 120 GACCCGAACGCGCGGCTCGGTGGTGACCACGAGAAATCCACCAAGGGTGCC---CTGAAG 176

Qy   56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
        |||||||| :: ||
Db 177 TTCTCTCCCTACTTGATGATCCCCAGACCCTGGGTACTACCAGTAC 236

Qy   76 ProAlaAspIleProAspTyrLysLysPheGlyGlyProLeu------GlyPheLysTrp 93
        |||
Db 237 CCG---GACGGAGACCCTCCGCTTCCAGACCCTGGGTATTGGAAGGATCGGGTATCAGTC 293

Qy   94 GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
             |||||| :::|||||| ||| |||||||||
Db 294 TACCGCGTGTTCGACTTTGAAGACGGAGGGGTCGTCACCGTTAAGTACTCCTAC 353

Qy 114 GlnAsnGlyCysPheIleTyrLysGlyValLysPheIleGlyValAsnPheProSerAspGly 133
        |||

ORIGIN

Alignment Scores:
```
Pred. No.:               2.07e-14      Length:        790
Score:                   209.50        Matches:        55
Percent Similarity:      47.25%        Conservative:   48
Best Local Similarity:   25.23%        Mismatches:    102
Query Match:             17.26%        Indels:         13
```

FEATURES   Location/Qualifiers
     source   1..790
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434P092"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="434 (synonym: htes3)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

COMMENT   This is the 5' sequence of the clone insert.
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No 81 sequence available.
            This clone (DKFZp434P092) is available at the RZPD in Berlin.
            Please contact the RZPD: Resourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            Contact: MIPS
            MIPS
JOURNAL   Unpublished (1999)
TITLE   EST (Ansorge, Benes, et al.)
AUTHORS   Ansorge,W., Benes,V., Krieger,S., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
REFERENCE   1 (bases 1 to 790)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
SOURCE   Homo sapiens (human)
KEYWORDS   EST.
VERSION   AL044652.1 GI:5432867
ACCESSION   AL044652
DEFINITION DKFZp434P092_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434P092 5', mRNA sequence.
LOCUS   AL044652   790 bp   mRNA   linear   EST 04-SEP-2003
RESULT 15
AL044652

```
Db   354 GAGGGGTTCCCATATCAAGGCCGAGCACATGAGACGTGATGGGAAGGCGGTTTCCTGACGACGGC 413
             |||      ::: |||  :::|||   ||||||
Qy   134 ProValMetGlnLeuLysLysThrMetGlyTyrProGluAlaSer---ThrGluArgLeuTyrPro 152
             |||||| :::  ||||| :::||||
Db   414 CCAGTCATGACCAGCCGAGATTGTCGACCAGGAGCGGCTCCGTGTCCAAGAAGACGTATCTT 473
Qy   153 ArgAspGlyValLeuLysGlyLeuHisSerLysAlaLeuLysLysAspGly----- 170
Db   474 AACGACAACACCATCGTGGACAGCTTCGACTGGAGTTACAACCTGCAGAANGGGAAGCGGC 533
Qy   171 -----GlyHisThrTyrLeuValGlu------PheLysSerIleTyrMet 182
Db   534 TACAGGGGCCCGAGTGTCGGAGCCACTACATCTCACATCAAAGGCCCTTTCAGCCGATCCATG 593
Qy   183 AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerIleLysLeuAspIleThr 202
Db   594 AAGAGACAGCCGGTG------TTCGTGTACCGGAAGTGCCACGTGAAG 635
Qy   203 SerHisAsnGluAspTyrThrIleValGluGln 213
Db   636 GCTTCCAAGACCGAAGTCACCCTGGACGAGAGG 668
```

DB: US-10-006-922a-12.p2n.rst

US-10-006-922A-12 (1-225) x AL044652 (1-790)
```
                         1                            6            Gaps:
```

```
Qy    16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGlyGluGly 35
Db    32 GTGGAACTGGATGGCCAATGGGACCATGAATGGGCACAAATTTCTGTCAGCGGGAGAGGGTGAAGGT 91
Qy    36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
Db    92 GATGCCACATACGGAAAGGCTCACCCTGAAATTCATCTGCACCACT---GGAAAGCTCCT 148
Qy    56 PheAlaTrpAspIleLeuSerProGlnProGlnThrTyrGlySerLysLysHis 75
Db   149 CTCCTCCCAGACAACCATTACCGTCTCCACCCAGTCTGGGCGTCTAAGAATCCCAACGAA 208
Qy    76 ProAlaAspIlePro------AspTyrLysGlyAspPheProGluPheLysTrp 93
Db   209 CCAGACCACATGAAGCAAGAGCAAGCCATGACTTTTTCAAGAGAGGGGTCTATGTGTCAG 268
Qy    94 GluArgMetAsnAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
Db   269 GAGAGGAACCATCTTTTTCAAAGATGACGGGGAACTACAAGACCCGCTGAAGTCAAGTTC 328
Qy   114 GlnAspGlyCysPheIleTyrLysValLysLysProPheIleGlyValAlaAsnPheProSerAspGly 133
Db   329 GAAGGTGGACACCCTGGTGAATAGAANGTGAGCTGAAGGCAGTTGACTTTAAGGAGGATGGA 388
Qy   134 ProValMetGlnLeuLysLysThrMetGlyTyrTrpGluAlaSerThrGluArgLeuTyrProArg 153
Db   389 AACATTCTCGGCCCACAAG---CTGGAATACAACTATAACTCCCACAATGTGTACATCATG 445
Qy   154 AspGlyValLeuLysGlyLeuHisSerLysAlaLeuLysLysAsp------Asp 169
Db   446 GCCGACAAGCAAAGAGAATGGGCAATCAAGGTCACTTCAAGATACAGACAACAACATTGAGGAT 505
Qy   170 GlyGlyHisThrTyrLeuValGluPhe------LysSerIleTyrMetAlaLysLysProVal 187
Db   506 GGATCCGTGCAGCTGGCCGGACCATTATCAACAGAACACTCCAATCGGCGACGGCGCCCTGTG 565
Qy   188 GlnIleLeuProGlyTyrTyrTyrValAspSerIleLysLeuAspIleThr------SerHis 204
Db   566 CTCCTCCCAGACAACCATTACCGTCTCCACCCAGTCTGCCCTGTCTAAGAATCCCAACGAA 625
Qy   205 AsnGluAspTyrThrIleValGluGlnIleTyrGluLysArgThrGluGlyArgHisHis 222
Db   626 AAGAGAGACCACATGGTCCTGCTGGAGTTTGTGACCGGCGTCGGGGATCACACAT 679
```

Search completed: January 12, 2006, 12:23:29
Job time : 3750 secs

THIS PAGE BLANK (USPTO)

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 12, 2006, 09:45:12 ; Search time 179 Seconds
         (without alignments)
         2234.365 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
                Xgapop 10.0 ; Xgapext 0.5
                Xgapop 10.0 ; Xgapext 0.5
                Pgapop   6.0 ; Pgapext 0.5
                Delop   6.0 ; Delext   7.0

Searched: 1303057 seqs, 888780828 residues
                Listing first 45 summaries

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Command line parameters:

Database : Issued_Patents_NA:*
  1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
  2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
  3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
  4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
  5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
  6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
  7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
  8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
  9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat_10012006_162409_18947/app_query.fasta_1.391
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCt=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_1_290 @runat_10012006_162409_18947 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	678	3	US-09-459-956-6	Sequence 6, Appli
2	1214	100.0	859	3	US-09-866-538-11	Sequence 11, Appl
3	1214	100.0	859	3	US-09-865-291-11	Sequence 11, Appl
4	1210	99.7	723	3	US-10-152-296-1	Sequence 1, Appli
5	1210	99.7	6984	3	US-10-001-189-45	Sequence 45, Appl
6	729.5	60.1	699	3	US-09-459-956-5	Sequence 5, Appli
7	681	56.1	860	3	US-09-244-779-1	Sequence 1, Appli
8	667	54.9	801	3	US-09-459-956-7	Sequence 7, Appli
9	495	40.8	690	3	US-09-459-956-2	Sequence 2, Appli
10	486.5	40.1	1079	3	US-09-609-161B-15	Sequence 15, Appl
11	486.5	40.1	1079	3	US-09-626-581D-64	Sequence 64, Appl
12	486.5	40.1	1079	3	US-09-415-765B-64	Sequence 64, Appl
13	486.5	40.1	1085	3	US-09-626-580C-64	Sequence 64, Appl
14	486.5	40.1	1085	3	US-09-277-716-15	Sequence 15, Appl
15	485.5	40.0	696	3	US-09-459-956-4	Sequence 4, Appli
16	485.5	40.0	720	3	US-10-652-703A-1	Sequence 1, Appli
17	485.5	40.0	720	3	US-10-652-703A-1	Sequence 1, Appli
18	485.5	40.0	720	3	US-10-021-818A-3	Sequence 3, Appli
19	477.5	39.3	696	3	US-09-839-650-2	Sequence 2, Appli
20	474.5	39.1	1021	3	US-09-459-956-3	Sequence 3, Appli
21	474.5	39.1	1021	3	US-10-652-703A-2	Sequence 2, Appli
22	474.5	39.1	1104	3	US-10-021-818A-3	Sequence 3, Appli
23	474.5	39.1	1104	3	US-09-609-161B-30	Sequence 30, Appl
24	474.5	39.1	1279	3	US-09-277-716-30	Sequence 30, Appl
25	474.5	39.1	1279	3	US-09-609-161B-31	Sequence 31, Appl
26	464.5	38.6	1482	3	US-10-021-818A-1	Sequence 1, Appli
27	464.5	38.0	720	3	US-10-021-818A-1	Sequence 1, Appli
28	218.5	18.0	1559	3	US-09-023-946B-3	Sequence 3, Appl
29	215.5	17.8	717	3	US-09-453-313-1	Sequence 1, Appli
30	215.5	17.8	4196	3	US-09-204-117B-1	Sequence 1, Appli
31	215.5	17.8	4199	3	US-09-204-117B-1	Sequence 1, Appli
32	215.5	17.8	6232	3	US-09-796-575-1	Sequence 1, Appli
33	215.5	17.8	4199	3	US-09-220-557-19	Sequence 19, Appl
34	215.5	17.8	7455	3	US-10-219-227-19	Sequence 19, Appl
35	215.5	17.8	7455	3	US-09-502-710-23	Sequence 23, Appl
36	215.5	17.8	7686	3	US-09-502-711-26	Sequence 26, Appl
37	215.5	17.8	7686	3	US-09-502-711-23	Sequence 23, Appl
38	215.5	17.8	7686	3	US-09-502-710-26	Sequence 26, Appl
39	215.5	17.8	7686	3	US-09-565-616A-2	Sequence 2, Appli
40	215.5	17.8	7687	3	US-09-502-710-24	Sequence 24, Appl
41	215.5	17.8	7687	3	US-09-502-711-24	Sequence 24, Appl
42	215.5	17.8	7687	3	US-08-818-604-31	Sequence 31, Appl
43	212.5	17.5	717	3	US-09-346-946-31	Sequence 21, Appl
44	212.5	17.5	717	3	US-09-023-946B-21	Sequence 21, Appl
45	212.5	17.5	764	2	US-08-818-604-30	Sequence 30, Appl

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-09-459-956-6
Sequence 6, Application US/09459956
Patent No. 6342379

GENERAL INFORMATION:
  APPLICANT: Tsien, Roger Y.
  APPLICANT: Gonzalez, III, Jesus E.
  TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
  TITLE OF INVENTION: OPTICAL METHODS
  FILE REFERENCE: REGEN1290-4
  CURRENT APPLICATION NUMBER: US/09/459,956
  CURRENT FILING DATE: 1999-12-13
  PRIOR APPLICATION NUMBER: 08/765,860
  PRIOR FILING DATE: 1999-05-08
  PRIOR APPLICATION NUMBER: 08/481,977
  PRIOR FILING DATE: 1995-06-07
  PRIOR APPLICATION NUMBER: PCT/US96/09652
  PRIOR FILING DATE: 1996-06-06
  NUMBER OF SEQ ID NOS: 22
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 6
  LENGTH: 678
  TYPE: DNA
  ORGANISM: Discosoma sp
  US-09-459-956-6

Alignment Scores:
  Pred. No.:   3.24e-160
  Score: 1214.00
  Percent Similarity: 100.00%
  Best Local Similarity: 100.00%
  Query Match: 100.00%
  DB: 3

  Length: 678
  Matches: 225
  Conservative: 0
  Mismatches: 0
  Indels: 0
  Gaps: 0

US-10-006-922A-12 (1-225) x US-09-459-956-6 (1-678)

```
Qy      1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db     61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGTCCATACGAAGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db    121 CACAATACCGTAAAGCTGAAAGTTACCAAGGGTGGCCCCTTGCCATTTGCTTGGGATATT 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db    181 TTGTCACCACAATTTCAGTATGGGAGCAAGGTGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy     81 AspTyrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db    241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db    301 GACGGTGGCGTCGTTACTGTGACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db    361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db    421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db    481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT 540

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db    541 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db    601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy    221 HisHisLeuPheLeu 225
Db    661 CACCATCTGTTCCTT 675
```

Alignment Scores:

```
Pred. No.:                          4.72e-160
Score:                              1214.00       Length:            859
Percent Similarity:                 100.00%       Matches:           225
Best Local Similarity:              100.00%       Conservative:       0
Query Match:                        100.00%       Mismatches:         0
DB:                                        3      Indels:             0
                                                  Gaps:               0
```

US-10-006-922A-12 (1-225) x US-09-866-538-11 (1-859)

```
Qy      1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 54

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db     54 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGTCCATACGAAGGC 113

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db    114 CACAATACCGTAAAGCTGAAAGTTACCAAGGGTGGCCCCTTGCCATTTGCTTGGGATATT 173

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db    174 TTGTCACCACAATTTCAGTATGGGAGCAAGGTGTATATGTCAAGCACCCTGCCGACATACCA 233

Qy     81 AspTyrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db    234 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 293

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db    294 GACGGTGGCGTCGTTACTGTGACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 353

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db    354 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 413

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db    414 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 473

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db    474 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT 533

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db    534 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 593

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db    594 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 653

Qy    221 HisHisLeuPheLeu 225
Db    654 CACCATCTGTTCCTT 713
```

```
; SOFTWARE: Patentin version 3.0
;
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
; DB: US-09-865-291-11

US-10-006-922A-12 (1-225) x US-09-865-291-11 (1-859)

; Alignment Scores:
; Pred. No.:
; Score:                4.72e-160
; Percent Similarity:   1214.00           Length:      859
; Best Local Similarity: 100.00%          Matches:     225
; Query Match:          100.00%           Conservative: 0
;                       3                 Mismatches:  0
;                                         Indels:      0
;                                         Gaps:        0

Qy      1 MetArgSerLysArgAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 ATGAGGTCTTCCAAGAGAATGTTAATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 113

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 ACGGTCAAATGGGCACCAGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 173

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy     81 AspThrTyrLysLeuSerPheProGlyGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    294 GACTACATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy    101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 GACGGTGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCGGTTATGCAAAAGAAGACA 473

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 533

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTCGGAT 653

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    654 ATAACAAGCCACAACGAAGACTATACAATTGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db    714 CACCATCTGTTCCTT 728
```

```
RESULT 4
US-10-152-296-1
; Sequence 1, Application US/10152296
; Patent No. 6723537
; GENERAL INFORMATION:
```

```
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-0001100US
; CURRENT APPLICATION NUMBER: US/10/152,296
; CURRENT FILING DATE:
; PRIOR APPLICATION NUMBER: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian
; OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
; OTHER INFORMATION: "red" red fluorescent protein (RFP)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(723)
; OTHER INFORMATION: DsRED
; US-10-152-296-1

US-10-006-922A-12 (1-225) x US-10-152-296-1 (1-723)

; Alignment Scores:
; Pred. No.:
; Score:                1.31e-159
; Percent Similarity:   1210.00           Length:      723
; Best Local Similarity: 100.00%          Matches:     224
; Query Match:          99.56%            Conservative: 1
;                       3                 Mismatches:  0
;                                         Indels:      0
;                                         Gaps:        0

Qy      1 MetArgSerLysArgAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4 GTGCGGCTCCTCCAAGAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 ACCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy     81 AspThrTyrLysLeuSerPheProGlyGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 GACTACAAGAAGCTGTCCTTCCCCGGAGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy    101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GACGGCGTGCGTGTGACCCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTCCTGAAGGGCGAG 483

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGTATC 543

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603
```

```
RESULT 5
US-10-001-189-45
  Sequence 45, Application US/10001189
; Patent No. 696810
; GENERAL INFORMATION:
; APPLICANT: LI, XU
; APPLICANT: BEAM, TERESA
; APPLICANT: FRASER JR., MALCOLM J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
; TITLE OF INVENTION: MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
; TITLE OF INVENTION: VECTOR PIGGYBAC
; FILE REFERENCE: 83591D-92098
; CURRENT APPLICATION NUMBER: US/10/001,189
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,984
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/244,677
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 45
; LENGTH: 6984
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: pBXP3-DsRed-orf sequence
US-10-001-189-45

Alignment Scores:
Pred. No.:
Score:                    4.87e-158
                          1210.00
Percent Similarity:       100.00%
Best Local Similarity:    99.56%
Query Match:              3

Length:        6984
Matches:       224
Conservative:  1
Mismatches:    0
Indels:        0
Gaps:          0

US-10-006-922A-12 (1-225) x US-10-001-189-45 (1-6984)
```

```
Qy      1 MetArgSerLysSerLysAsnValIleLysGluMetArgPheLysValArgMetGluGly 20
Db   2974 GTGCCGCTCCTCCCAAGAGACGTCATCAAGGAGTTCAGGGTGCGCATGGAGGGC 3033

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   3034 ACCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCTACGAGGGC 3093

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   3094 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3153

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   3154 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3213

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   3214 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3273

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   3274 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3333

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   3334 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 3393

Qy    161 IleHisHisLeuPheLeu 225
Db    664 CACCACCTGTTCCTG 678
```

```
Qy    201 IleThrSerHisAsnGluAspThrTyrThrIleValGlnGluGlnThrGluGlyGlyArg 220
Db    604 ATCACCTCCCACAACGAGGACACCTACACCATCGTGCAGGAGCAGACCGAGGGCGGC 663

Qy    221 HisHisLeuPheLeu 225
Db    664 CACCACCTGTTCCTG 678
```

```
RESULT 6
US-09-459-956-5
  Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

Alignment Scores:
Pred. No.:
Score:                    2.87e-92
                          729.50
Percent Similarity:       76.00%
Best Local Similarity:    59.11%
Query Match:              3

Length:        699
Matches:       133
Conservative:  38
Mismatches:    53
Indels:        1
Gaps:          1

US-10-006-922A-12 (1-225) x US-09-459-956-5 (1-699)
```

```
Qy      1 MetArgSerLysSerLysAsnValIleLysGluMetArgPheLysValArgMetGluGly 20
Db      1 ATGAGGTCGTTCCAAGAGTGTGATCAAGGAGAAGAAAATGTTGATCGATCTTCATCTGGAAGGA 60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db     61 ACGGTTCAATGGGCACTACTTTGAAATAAAAGGCAAAAGGACAGCCTTAATGAAGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db    121 ACCAATACCGTCACGCCTCGAGGTTACCAAGGGGTGGACCTCTGCCCAATTGGTTGGCATATT 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db        ...
```

```
Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   3454 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3513

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   3514 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3573

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   3574 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 3633

Qy    201 IleThrSerHisAsnGluAspThrTyrThrIleValGlnGluGlnThrGluGlyGlyArg 220
Db   3634 ATCACCTCCCACAACGAGGACACCTACACCATCGTGCAGGAGCAGACCGAGGGCGGC 3693

Qy    221 HisHisLeuPheLeu 225
Db   3634 CACCACCTGTTCCTG 3648
```

RESULT 7
US-10-244-779-1

; Sequence 1, Application US/10244779
; Patent No. 6933375
; GENERAL INFORMATION:
; APPLICANT: Falkowski, Paul
; APPLICANT: Sun, Yi
; APPLICANT: Gorbunov, Maxim
; APPLICANT: Wyman, Kevin
; APPLICANT: Chen, Yi-Bu
; TITLE OF INVENTION: mcFP Encoding Nucleic Acids,
; TITLE OF INVENTION: Polypeptides, Antibodies and Methods of Use Thereof
; FILE REFERENCE: Rut 00-0023US
; CURRENT APPLICATION NUMBER: US/10/244,779
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: 60/322,189
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 860
; TYPE: DNA
; ORGANISM: Montastrea cavernosa
US-10-244-779-1

Alignment Scores:

Pred. No.:
Score: 2.52e-85
Percent Similarity: 73.95%
Best Local Similarity: 57.21%
Query Match: 56.10%
DB: 3

Length:	860
Matches:	123
Conservative:	36
Mismatches:	56
Indels:	0
Gaps:	0

```
Db  181 TTGTGCCCACAATTTCAGTATTGGAAACAACAGGCATTTGTCCACCACCCTGACAACATACAT 240
Qy   81 AspTyrLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
Db  241 GATTATCTAAAGCTGTTCATTTCCGGAGGATATACATGGGAACGGTCCCATGCACTTTGAA  300
Qy  101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
Db  301 GACGGTGGCGTTGCTGTTGTTGTATCACCAATGATATCAGTTGACAAGGCAACTGTTTCTACTAC 360
Qy  121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 GACATCAAGTTCACTGGCGTTGAACTTTCCTCCAAATGGACACCGGTTGTCAGAAGAAGACA 420
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlu 160
Db  421 ACTGGCTGGGAAGCCAGCACTGAGCGTCTTTGTATCCTCCGTGACGGTGTTGTTGACAGAGAC 480
Qy  161 IleHisLysLysAlaLeuLysLeuAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
Db  481 ATCCATAAGTTCTGACAGTTGAAGAGGTGGTCATTACGAATGACACATTAAAAACTGTT   540
Qy  181 TyrMetAlaLysLys---ProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeu  199
Db  541 TACAGGGGCCAAGAAGGCCGCCGTTGAGATGCCAGGGTTATCCATATGTTGACACCGAAACTG 600
Qy  200 AspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGly   219
Db  601 GTTATATGGGAACAACGACAAAAGAATTCATGAAAGTTGAGGAGCCAATGAAATGCCGTTGCA 660
Qy  220 ArgHisHisLeuPhe 224
Db  661 CGCCACCATCCGTTC 675
```

US-10-006-922A-12 (1-225) x US-10-244-779-1 (1-860)

```
Qy    6 AsnValIleLysGluPheLysValAlaArgMetGlyThrValAsnGlyHis  25
Db  138 AGTGTGATTAAATCAGTCATGAAATCAAGCTGCGTATGGACGGGCATTGTAAACGGGCAC 197
```

RESULT 8
US-09-459-956-7

; Sequence 7, Application US/09459956
; Patent No. 6343379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavularia sp
US-09-459-956-7

Alignment Scores:

Pred. No.:
Score: 2.06e-83
Percent Similarity: 73.76%
Best Local Similarity: 54.75%
Query Match: 54.94%
DB: 3

Length:	801
Matches:	121
Conservative:	42
Mismatches:	58
Indels:	0
Gaps:	0

```
Qy   26 GluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLys  45
Db  198 AAGTTCATGATTACAGGAGAAGGTGAAGGCAAGGCCTTTGAGGGAACACACTATATA   257
Qy   46 LeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPhe  65
Db  258 CTTAAAGTCAGAAGAGGCGGACCTCTGCCTTTCGCTTGGGACATCTTGACAGCCAGCATTT 317
Qy   66 GlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLysLeu  85
Db  318 CAGTACGGCAGCAAGGTTTACGTCAAAAATACACCAGCAGACTATTCAAGCAG   377
Qy   86 SerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValVal  105
Db  378 TCGTTTCCTGAGGGTTTCAAGTGGGAAAGAGTAAGCATGAAAGACGGCGGTGTTGTTGTA  437
Qy  106 ThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIle  125
Db  438 ACCGTCACAGCGACCATAAAGTTGGAAGGCAGCATGTTTTTCTACGAAAGTTGGATTTAT  497
Qy  126 GlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAla 145
Db  498 GGTGTGAACTTTCCCTCCAGTGGTCCAGTTATGCAGAAGAAGACGGCTGAAATGGGAGCCA 557
Qy  146 SerThrGluArgLeuTyrProArgAspGlyValLeuLeuGluIleHisLysAlaLeu  165
Db  558 TCCACTGAGAATATGTACGTGCGTGATGGAGTGCTACTGGTTGAGCAGCTG  617
Qy  166 LysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys  185
Db  618 TTGCTTGAAGGGGATAAACATCACCGAGTAACTTCAGAAGTACTTACGGGCGAAAGAAG  677
Qy  186 ProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuIleThrSerHisAsn  205
Db  678 GGTGTGCGTGTTGCCAGAATATCACTTCGTGGACCAACCGAATTGAAAATTCTGAGCCATGAC 737
Qy  206 GluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  738 AAAGATTACAACACCGTTGAGGTGTATGAGAAGTGCCGTTGCTGC  782
```

US-10-006-922A-12 (1-225) x US-09-459-956-7 (1-801)

```
Qy      3 SerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
            ||||:||
Db    121 ACCACCATGGGTGTGATTAAACCAGAACATGAAGACTTGAAGATTAAGCTGAAGAGGAAGAAAATGTA 180

Qy     23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 42
Db     23 AACGCGTTCCAGTACGGGAAACAGAGCATTGACAAAAATACCCAGACGATATAGCACAGACTAT 180

...
```

Alignment Scores:

Pred. No.:		
Score:	2.11e-59	
	495.00	Length: 690
Percent Similarity:	64.68%	Matches: 92
		Conservative: 38
Best Local Similarity:	45.77%	Mismatches: 69
Query Match:	40.77%	Indels: 2
DB:	3	Gaps: 1

US-10-006-922A-12 (1-225) x US-09-459-956-2 (1-690)

```
Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
            ||| :
Db      1 ATGGCGTCTTCAAACAGGTTTATCAGAAGTTTATCGGAGATGACAGAAATGACCTTACCAATATGGATGGC 60
```

Left column

```
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/102,939
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/089,367
PRIOR FILING DATE: 1998-06-15
PRIOR APPLICATION NUMBER: 60/079,624
PRIOR FILING DATE: 1998-03-27
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 15
LENGTH: 1079
TYPE: DNA
ORGANISM: Renilla mulleri
FEATURE:
NAME/KEY: CDS
LOCATION: (259)..(975)
OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-15
```

```
Alignment Scores:
Pred. No.:
Score:                   6.71e-58        Length:       1079
Percent Similarity:      486.50          Matches:      97
Best Local Similarity:   64.09%          Conservative: 44
Query Match:             44.09%          Mismatches:   72
DB:                      40.07%          Indels:       7
                         3               Gaps:         3

US-10-006-922A-12 (1-225) x US-09-609-161B-15 (1-1079)
```

```
Qy    2   ArgSerSerLysAsnValIleLys------------------GluPheMetArgPheLysVal  16
              ::|   |||||                               ||:||||||||||
Db  256   AAGATGAGTAAACAAAAATATTGAAGAAGACACTTGTTTACAAGAGTAAATGTCGTATAAAGTA  315

Qy   17   ArgMetGluGlyThrValAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyArg  36
              :::|   ||||||                    ::|||||||||||||||||||||
Db  316   AATCTGGAAGGAATTGTAAACAACCATGTTTTTTACAANNGGAGGGTTGCGGCAAAGGGAAT  375

Qy   37   ProTyrGluGlyHisAsnThrValLysLeuLysGlyGlyProLeuProPhe  56
              |||||   |||||||                  ::|||||||||||||||
Db  376   ATTTTATTCGGCAATCAACTGGTTCAGATTCGTGTGTCACCGAAAAGGGCCCCACTGCCTTTT  435

Qy   57   AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisPro  76
              |||||   ||||||||                  ::|||||||||||||||||||
Db  436   GCATTTGATATTGTCTCACCAGCTTTTCAAATATGGCAACCGTACTTCTCACGAAAATATCCG  495

Qy   77   AlaAspIleProAspTyrLysLysValAlaAsnPheProSerArgAspGlyProValMet  96
              |||||   ||||                    ::|||||||||||||
Db  496   AATGATATATCAGATTATTTTATACAATCATTTCCAGCAGGATTTAGTGAACGAACA  555

Qy   97   MetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGly  116
              |||||   ||||                     ::|||||||||||
Db  556   TTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAGATATAAAATTTAATAGAAGAC  615

Qy  117   CysPheIleTyrLysValLysPheIleGlyValAlaAsnPheProSerArgAspGlyProValMet  136
              ||||    |||||                    ::|||||||||||
Db  616   AAGTTCGTCTACAGAGTGGAATACAAAAGGTAGTGAACTCCCAGATGATGAGTCCGTCATG  675

Qy  137   GlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyVal  156
              |||||   |||||                    ::|||||||||||
Db  676   CAGAAGACTATCTTAGGAATAGAGCCTTCATTGGAAGCCAATGTACAATGAATAATGGCGTC  735

Qy  157   LeuLysGlyGlyIleLeuHisIleLysSerAlaLeuLysLysAspGlyGlyHisTyrLeuValGlu  176
              |||||   ||||                    ::|||||||||||
Db  736   TTGGTCGGCGAAGTAATTCTTGTCTTTAAAACTCTGGGAAATAATAATTAATTCGATGTCAC  795

Qy  177   PheLysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrTyrVal  195
              |||||   ||||                    :::|||||||||||
Db  796   ATGAAAACATTAATGAAGTCGAAAGGGTGTTAAAGGAGTTTCCTTCTTCGTATCATTTATT  855

Qy  196   AspSerLysGlyLeuAspIleThrSerHisAsnGluAspTyrThrIleValGlnLnTyrGlu  215
              :::|||   |||||                    ::|||||||||||
Db  856   CAACATCGTTTGGAAAAGACT---TACGTAAGAAGACGGGGGGGGTTCGTTGAACAGGCATGAG  912
```

Right column

```
us-10-006-922a-12.p2n.rnl

RESULT 11
US-09-626-581D-64
Sequence 64, Application US/09626581D
Patent No. 6548249
GENERAL INFORMATION:
APPLICANT: Anderson, David
TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
TITLE OF INVENTION: Libraries
FILE REFERENCE: A-66900-3/RMS
CURRENT APPLICATION NUMBER: US/09/626,581D
CURRENT FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 09/169,015
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 09/415,765
PRIOR FILING DATE: 1999-10-08
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 64
LENGTH: 1079
TYPE: DNA
ORGANISM: Renilla mulleri
FEATURE:
NAME/KEY: CDS
LOCATION: (259)..(975)
OTHER INFORMATION:
US-09-626-581D-64
```

```
Alignment Scores:
Pred. No.:
Score:                   6.71e-58        Length:       1079
Percent Similarity:      486.50          Matches:      97
Best Local Similarity:   64.09%          Conservative: 44
Query Match:             44.09%          Mismatches:   72
DB:                      40.07%          Indels:       7
                         3               Gaps:         3

US-10-006-922A-12 (1-225) x US-09-626-581D-64 (1-1079)
```

```
Qy    2   ArgSerSerLysAsnValIleLys------------------GluPheMetArgPheLysVal  16
              ::|   |||||                               ||:||||||||||
Db  256   AAGATGAGTAAACAAAAATATTGAAGAAGACACTTGTTTACAAGAGTAAATGTCGTATAAAGTA  315

Qy   17   ArgMetGluGlyThrValAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyArg  36
              :::|   ||||||                    ::|||||||||||||||||||||
Db  316   AATCTGGAAGGAATTGTAAACAACCATGTTTTTTACAANNGGAGGGTTGCGGCAAAGGGAAT  375

Qy   37   ProTyrGluGlyHisAsnThrValLysLeuLysGlyGlyProLeuProPhe  56
              |||||   |||||||                  ::|||||||||||||||
Db  376   ATTTTATTCGGCAATCAACTGGTTCAGATTCGTGTGTCACCGAAAAGGGCCCCACTGCCTTTT  435

Qy   57   AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisPro  76
              |||||   ||||||||                  ::|||||||||||||||||||
Db  436   GCATTTGATATTGTCTCACCAGCTTTTCAAATATGGCAACCGTACTTCTCACGAAAATATCCG  495

Qy   77   AlaAspIleProAspTyrLysLysValAlaAsnPheProSerPheProGluGlyProLeuProPhe  96
              |||||   |||||                    ::|||||||||||||
Db  496   AATGATATATCAGATTATTTTATACAATCATTTCCAGCAGGATTTAGTGATGAACGAACA  555

Qy   97   MetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGly  116
              |||||   ||||                     ::|||||||||||
Db  556   TTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAGATATAAAATTTAATAGAAGAC  615

Qy  117   CysPheIleTyrLysValLysPheIleGlyValAlaAsnPheProSerArgAspGlyProValMet  136
              ||||    |||||                    ::|||||||||||
Db  616   AAGTTCGTCTACAGAGTGGAATACAAAAGGTAGTGAACTCCCAGATGATGAGTCCGTCATG  675

Qy  137   GlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyVal  156
              |||||   |||||                    ::|||||||||||
Db  676   CAGAAGACTATCTTAGGAATAGAGCCTTCATTGGAAGCCAATGTACAATGAATAATGGCGTC  735

Qy  157   LeuLysGlyGlyIleLeuHisIleLysSerAlaLeuLysLysAspGlyGlyHisTyrLeuValGlu  176
              |||||   ||||                    ::|||||||||||
Db  736   TTGGTCGGCGAAGTAATTCTTGTCTTTAAAACTAAACTCTGGGAAATAATAATTAATTCGATGTCAC  795
```

```
Qy  177  PheLysSerIleTyrMetAlaLeuLysLysProVal---GlnLeuProGlyTyrTyrTyrVal  195
         |||||||||||||||||||||||||||     :::|||||||||||||||||||||||
Db  796  ATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGTTTCCTTCGTATCATTTTATT  855

Qy  196  AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu  215
         ::||||        ||||       ||||||||
Db  856  CAACATCGTTTGGAAAAGACT---TACGTAGAAGACGGGGGGGGTTCGTTGAACAGCATGAG  912
```

RESULT 12
US-09-415-765B-64
; Sequence 64, Application US/09415765B
; Patent No. 648632
; GENERAL INFORMATION:
;   APPLICANT: Anderson, David
;   TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
;   TITLE OF INVENTION: Libraries
;   FILE REFERENCE: A66900-1/RMS/AMS
;   CURRENT APPLICATION NUMBER: US/09/415,765B
;   CURRENT FILING DATE: 1999-10-08
;   PRIOR APPLICATION NUMBER: 09/169,015
;   PRIOR FILING DATE: 1998-10-08
;   NUMBER OF SEQ ID NOS: 65
;   SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   OTHER INFORMATION:
US-09-415-765B-64

Alignment Scores:
Pred. No.:
Score: 6.71e-58
                         486.50   Length: 1079
Percent Similarity: 64.09%   Matches: 97
Best Local Similarity: 44.09%   Conservative: 44
Query Match: 40.07%   Mismatches: 72
DB: Indels: 7
                                      Gaps: 3

US-10-006-922A-12 (1-225) x US-09-415-765B-64 (1-1079)

```
Qy    2  ArgSerLysSerLysAsnValIleLys------------GluPheMetArgPheLysVal  16
           ::       |||||        :::|||||      ::   ||
Db  256  AAGAGTAGGTAAACAAATATGTGAAGAACACTTGTTTTACAGAAGTAATGTCGTATAAAGTA  315

Qy   17  ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArg  36
           :::|||        |||||     ||||||||
Db  316  AATCTGGAAGGAAATTGTAAACAACCATGTTTTACATGGAGGGGTTGCGGCAAAGGGAAT  375

Qy   37  ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPhe  56
         |||||||       |||||
Db  376  ATTTTATTCGGCAATCAACTGGTTCAGATTCGGTGTCACGAAAAGGGCCCCACTGCCTTTT  435

Qy   57  AlaLeuTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValThrValLysHisPro  76
          |||     ||||        |||::::||||||||||||||||     :::|||
Db  436  GCATTTGATTGTGTCCAGCTTTTCAATATGGCAAGCCTACTTCACGAAAAATATCCG  495

Qy   77  AlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgVal  96
         |||||       |||||
Db  496  AATGATATCAGATTTATAAAGAAATTTATGATGTACTTCCAGCAGGATTTATGTAATGAACGAACA  555

Qy   97  MetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGly  116
          |||||||||||
Db  556  TTACGTTACGGAAGATGGCGGACTTGTTGAAATTCGTTCAGATATAAATTTAATAGAAGAC  615

Qy  117  CysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMet  136
         ||||||       ::|||
Db  616  AAGTTCGTCTACAGACGTGGAATAACTTCCAGAGTGAGTAACTTCAAAAGGGTGTCCCTCATG  675

Qy  137  GlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyVal  156
         ::||||||       |||
```

```
Db  676  CAGAGACTATCTTAGGAATAGAGCCTTCATTTGAGTACATGAATAAATGCGTC  735

Qy  157  LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrIleValGlu  176
         :::|||||       ::||||
Db  736  TTGGTCGGCGGAAGTAATTCTTGTCTCTATAAAAACCTGGGAAATATTATTCAGTCAC  795

Qy  177  PheLysSerIleTyrMetAlaLeuLysLysProVal---GlnLeuProGlyTyrTyrTyrVal  195
         ::||||||||||||||||||||||||||||     :::||||||||||||||
Db  796  ATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGTTTCCTTCGTATCATTTTATT  855

Qy  196  AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu  215
         ::||||       :::||||
Db  856  CAACATCGTTTGGAAAAGACT---TACGTAGAAGACGGGGGGGGTTCGTTGAACAGCATGAG  912
```

RESULT 13
US-09-626-580C-64
; Sequence 64, Application US/09626580C
; Patent No. 6552617
; GENERAL INFORMATION:
;   APPLICANT: Anderson, David
;   APPLICANT: Bogenberger, Jakob M.
;   APPLICANT: Peele, Beau R.
;   TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
;   FILE REFERENCE: A-66900-2/RMS/AMS
;   CURRENT APPLICATION NUMBER: US/09/626,580C
;   CURRENT FILING DATE: 2000-07-27
;   PRIOR APPLICATION NUMBER: US 09/415,765
;   PRIOR FILING DATE: 1999-10-08
;   PRIOR APPLICATION NUMBER: US 09/169,015
;   PRIOR FILING DATE: 1998-10-08
;   NUMBER OF SEQ ID NOS: 65
;   SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   OTHER INFORMATION:
US-09-626-580C-64

Alignment Scores:
Pred. No.:
Score: 6.71e-58
                         486.50   Length: 1079
Percent Similarity: 64.09%   Matches: 97
Best Local Similarity: 44.09%   Conservative: 44
Query Match: 40.07%   Mismatches: 72
DB: Indels: 7
                                      Gaps: 3

US-10-006-922A-12 (1-225) x US-09-626-580C-64 (1-1079)

```
Qy    2  ArgSerLysSerLysAsnValIleLys------------GluPheMetArgPheLysVal  16
           ::       |||||        :::|||||      ::   ||
Db  256  AAGAGTAGGTAAACAAATATGTGAAGAAC...                               315

Qy   17  ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArg  36
           :::|||        |||||     ||||||||
Db  316  AATCTGGAAGGAAATTGTAAACAACCATGTTTTACATGGAGGGGTTGCGGCAAAGGGAAT  375

Qy   37  ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPhe  56
         |||||||       |||||
Db  376  ATTTTATTCGGCAATCAACTGGTTCAGATTCGGTGTCACGAAAAGGGCCCCACTGCCTTTT  435

Qy   57  AlaLeuTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValThrValLysHisPro  76
          |||     ||||        |||::::||||||||||||||||     :::|||
Db  436  GCATTTGATTGTGTCCAGCTTTTCAATATGGCAAGCCTACTTCACGAAAAATATCCG  495

Qy   77  AlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgVal  96
         |||||       |||||
Db  496  AATGATATCAGATTTATAAAGAAATTTATGATGTACTTCCAGCAGGATTTATGTAATGAACGAACA  555

Qy   97  MetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGly  116
          |||||||||||
```

RESULT 14
US-09-277-716-15

; Sequence 15, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE L
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-15

Alignment Scores:
Score: 6.77e-58
Pred. No.:   3
Percent Similarity: 40.07%
Best Local Similarity: 44.09%
Query Match: 3

US-10-006-922A-12 (1-225) x US-09-277-716-15 (1-1085)

Length: 1085
Matches: 97
Conservative: 44
Mismatches: 72
Indels: 7
Gaps: 3

RESULT 15
US-09-459-956-4

; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp

US-09-459-956-4

Alignment Scores:
Score: 4.6e-58
Pred. No.:   3
Percent Similarity: 64.71%
Best Local Similarity: 46.57%
Query Match: 3

US-10-006-922A-12 (1-225) x US-09-459-956-4 (1-696)

Length: 696
Matches: 95
Conservative: 37
Mismatches: 67
Indels: 5
Gaps: 3

```
Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGlyGluGlyGlyGluGlyArgProTyrGluGly   40
             ||||||||||||||||||||:::|||       |||||||||||||
Db   61  TGCGTCAACCGGACATAAATTTGTGATCACGGGCGGGGAAGGCAATTGGAATATCCGTTCAAAGGG  120

Qy   41  HisAsnThrValLysLeuValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
             |||   |||||   |||   ||| |::||||   |||   ||:::
Db  121  AAACAGACTATTAAATCTGTGTGATCGAAGGGGGACCCATTGCCATTTTCCGAAGACATA  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
             |||||||||||||   ||||||||   |||   ||||||||||||   |||
Db  181  TTGTCAGCTGGCTTTAAGTACGGAGAGACAGGATTTTCACTGAATAATCCTCAAGACATAGTA  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
             ||||||||||||   ||| |||   |||   ||||||||||||   ||||||
Db  241  GACTATTTCAAGAACTCGTGTCCTGCTGGATATACATGGGCCAGGTCTTTTCTCTTTGAG  300

Qy  101  AspGlyGlyValValThrValThrGlnAsp------SerSerLeuGlnAspGlyCysPhe  118
             |||||||||   |   |||   ||||   |||:::::::::
Db  301  GATGGAGCAGTCTCGCATATGCAAATGTAGATATTAACAGTGAGTGTCAAAAGAAAACTGCATT  360

Qy  119  IleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLys  138
             |||   |:::||||||   |||   ||||   ||||::::::
Db  361  TATCAATAAGAGCATATTTAATGGAATGAATTTTCCTGCTGAAGGACCTGTCGATGAAAAG  420

Qy  139  LysThrMetGlyTrpGluAlaSerThrGluArgLeu------TyrProArgAspGlyVal  156
             |||||||||||||||||   |:::::::   :::
Db  421  ATGACAACTAACTGGGAAGCATCCTGCGAAGAATCATGCCAGTACCTAAGCACAGGGATA  480

Qy  157  LeuLysGlyGluIleHisLysAlaLeuLysLeuAspGlyGlyHisTyrLeuValGlu  176
             ||||||||||||||||||||   |||||||||||||||   :::
Db  481  CTGAAAAGGGGATGTCTTCCATGTACCTCCTTCTTGAAGGATGGTGGGGGGTTACGGGTGCCAG  540

Qy  177  PheLysSerIleTyrMetAlaLysLys---ProValGlnLeuProGlyTyrTyrTyrVal  195
             :::::::   |||||||||||   |||||||   |||   ||||||||||
Db  541  TTCGACACAGTTTACAAAGCAAAAGTCTGTGCCAAGTAAGAATGCCGAGTGGCACTTCATC  600

Qy  196  AspSerLysLeu  199
             ||||||
Db  601  CAGCATAAGCTC  612
```

Search completed: January 12, 2006, 12:26:40
Job time : 185 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 12, 2006, 10:08:26 ; Search time 793 Seconds
(without alignments)
2346.293 Million cell updates/sec

Title: US-10-006-922A-12
Perfect Score: 1214
Sequence: 1 MRSSKNVIKEFMRKVRMEG...........EDYTIVEQYBRTEGRHHLFL 225

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:

MODEL=frame+_p2n.model -DEV-xlp
-Q=/cgn2_1/USPTO_spool_D/US10006922/runat 10012006 162409 18977/app_query.fasta_1.391
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN 1 1 1549 @runat 10012006 162409 18977 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	678	3	US-09-967-772-6	Sequence 6, Appli
2	1214	100.0	678	5	US-10-006-922-11	Sequence 11, Appl
3	1214	100.0	678	7	US-10-081-864-7	Sequence 7, Appli
4	1214	100.0	678	5	US-10-121-258-2	Sequence 2, Appli
5	1214	100.0	678	6	US-10-315-920-1	Sequence 1, Appli
6	1214	100.0	678	6	US-10-132-067-3	Sequence 3, Appli
7	1214	100.0	678	6	US-10-335-517-6	Sequence 6, Appli
8	1214	100.0	678	6	US-10-334-288-6	Sequence 6, Appli
9	1214	100.0	678	7	US-10-311-030-5	Sequence 5, Appli
10	1214	100.0	678	8	US-10-656-029-21	Sequence 21, Appl
11	1214	100.0	678	9	US-10-505-486-27	Sequence 27, Appl
12	1214	100.0	678	6	US-10-844-064A-1	Sequence 1, Appli
13	1214	100.0	678	9	US-09-931-304-2	Sequence 2, Appli
14	1214	100.0	859	9	US-09-999-745-66	Sequence 66, Appl
15	1214	100.0	859	3	US-09-866-538-11	Sequence 11, Appl
16	1214	100.0	859	3	US-09-794-308-11	Sequence 11, Appl
17	1214	100.0	859	3	US-09-865-291-11	Sequence 11, Appl
18	1214	100.0	859	8	US-10-885-988-11	Sequence 11, Appl
19	1214	100.0	859	8	US-10-931-304-23	Sequence 23, Appl
20	1214	100.0	859	8	US-09-857-622-11	Sequence 11, Appl
21	1214	100.0	3311	8	US-10-311-030-12	Sequence 12, Appl
22	1210	99.7	681	5	US-09-797-496B-3	Sequence 3, Appli
23	1210	99.7	681	5	US-10-006-922-35	Sequence 35, Appl
24	1210	99.7	681	7	US-10-121-258-3	Sequence 3, Appli
25	1210	99.7	681	7	US-10-121-258-23	Sequence 23, Appl
26	1210	99.7	681	9	US-10-931-304-3	Sequence 3, Appli
27	1210	99.7	681	9	US-10-931-304-23	Sequence 23, Appl
28	1210	99.7	713	9	US-10-311-030-11	Sequence 11, Appl
29	1210	99.7	713	7	US-10-311-030-12	Sequence 12, Appl
30	1210	99.7	723	7	US-10-152-296-1	Sequence 1, Appli
31	1210	99.7	723	7	US-10-739-656-1	Sequence 1, Appli
32	1210	99.7	1638	5	US-10-214-932-51	Sequence 51, Appl
33	1210	99.7	1647	5	US-10-214-932-75	Sequence 75, Appl
34	1210	99.7	4692	6	US-10-161-403-29	Sequence 29, Appl
35	1210	99.7	4692	7	US-10-433-640-16	Sequence 16, Appl
36	1210	99.7	4692	10	US-11-006-076-29	Sequence 29, Appl
37	1210	99.7	5436	5	US-10-169-050-46	Sequence 46, Appl
38	1210	99.7	6984	5	US-10-001-189-45	Sequence 45, Appl
39	1210	99.7	7910	6	US-10-169-050-20	Sequence 20, Appl
40	1210	99.7	9320	8	US-10-471-065-20	Sequence 20, Appl
41	1210	99.7	9658	8	US-10-609-019-4	Sequence 4, Appli
42	1210	99.7	9678	8	US-10-609-019-3	Sequence 3, Appli
43	1210	99.7	10263	8	US-10-609-019-2	Sequence 2, Appli
44	1207	99.4	678	6	US-10-315-920-3	Sequence 3, Appli
45	1203	99.1	7508	7	US-10-742-828-4	Sequence 4, Appli

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-09-967-772-6
; Sequence 6, Application US/09967772
; Patent No. US20020164577A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/09/967,772
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 08/765,860
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT/ US96/09652
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: US 08/481,977
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
US-09-967-772-6

Alignment Scores:
Pred. No.:   6.83e-143   Length: 678
Score: 1214.00   Matches: 225

```
Percent Similarity:        100.00%        Conservative:  0
Best Local Similarity:     100.00%        Mismatches:    0
Query Match:               100.00%        Indels:        0
DB:                        3              Gaps:          0

US-10-006-922A-12 (1-225) x US-09-967-772-6 (1-678)

Qy    1 MetArgSerSerLysAsnVallIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1 ATGAGGTCTTCCAAGAAATGTTATCAGAAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGGACCTTTGCCATTTGCTTGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCAATTTCAGTATGCAAGGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPhelysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGAAAGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGGGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleGlnGlnHisAlaValAlaArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGATACAGCAGCATGCAGTAGCAAGGACTGAGGGACGC 660

Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

```
Alignment Scores:
Pred. No.:                 6.83e-143
Score:                     1214.00       Length:        678
Percent Similarity:        100.00%       Matches:       225
Best Local Similarity:     100.00%       Conservative:  0
Query Match:               100.00%       Mismatches:    0
DB:                        5             Indels:        0
                                         Gaps:          0

US-10-006-922A-12 (1-225) x US-10-006-922-11 (1-678)

Qy    1 MetArgSerSerLysAsnVallIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1 ATGAGGTCTTCCAAGAAATGTTATCAGAAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGGACCTTTGCCATTTGCTTGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCAATTTCAGTATGCAAGGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPhelysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGAAAGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGGGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleGlnGlnHisAlaValAlaArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGATACAGCAGCATGCAGTAGCAAGGACTGAGGGACGC 660

Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT
```

RESULT 3
US-10-081-864-7
; Sequence 7, Application US20030022287A1
; Publication No. US20030022287A1 Aggregating Fluorescent Proteins and
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savistky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: Methods for Using the Same
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
; DB:

Alignment Scores:
Pred. No.:
Score: 1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 5   Mismatches: 0
                                         Indels: 0
                                         Gaps: 0

US-10-006-922A-12 (1-225) x US-10-081-864-7 (1-678)

Qy   1 MetArgSerLysAsnValIleLeuLysGlyHisGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGCGGCTCTCCAAGAACGTTCATCAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCTTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACCTGTTCCTT 675

RESULT 4
US-10-121-258-2
; Sequence 2, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(678)
; OTHER INFORMATION: wild-type DsRed
US-10-121-258-2

Alignment Scores:
Pred. No.:
Score: 1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 5   Mismatches: 0
                                         Indels: 0
                                         Gaps: 0

US-10-006-922A-12 (1-225) x US-10-121-258-2 (1-678)

Qy   1 MetArgSerLysAsnValIleLeuLysGlyHisGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGCGCTTCAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACCGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCAGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGCTTAAGGTTACCAAGGGTGGCCCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 121 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

RESULT 5

US-10-315-920-1

; Sequence 1, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Terskikh, Alexey
; APPLICANT: Pradkov, Arcady Fedorovich
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-1

Alignment Scores:
Pred. No.:
Score: 6.83e-143
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 6   Indels: 0
                                          Gaps: 0

US-10-006-922A-12 (1-225) x US-10-315-920-1 (1-678)

RESULT 6

US-10-132-067-3

; Sequence 3, Application US/10132067
; Publication No. US20020203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands with Intrinsic
; TITLE OF INVENTION: Fluorescence
; FILE REFERENCE: 021362-00060005
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; OTHER INFORMATION: red fluorescent protein (dsRED)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(678)
; OTHER INFORMATION: dsRED
US-10-132-067-3

Alignment Scores:
Pred. No.:
Score: 6.83e-143
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 6   Indels: 0
                                          Gaps: 0

```
US-10-006-922A-12 (1-225) x US-10-132-067-3 (1-678)

Qy    1 MetArgSerGlyHisGluPheGlyIleGluGlyMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGCCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACCGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAATACCGTAAAGCTTAAAGGTAACCAAGGGGGCCCTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTGTCACCACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGluIleTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGGAGATCTATGAAAGAACCGAGGGACGC 660

Qy  221 HisHisHisLeuPheLeu 225
        |||||||||||||||||
Db  661 CACCATCTGTTCCTT 675
```

RESULT 7
US-10-335-517-6

; Sequence 6, Application US/10335517
; Publication No. US20030207248A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/10/335,517
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/967,772
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 08/765,860
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT/ US96/09652
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: US 08/481,977
; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
US-10-335-517-6

Alignment Scores:
Pred. No.:
Score: 1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:
DB:

Length: 678
Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-335-517-6 (1-678)

```
Qy    1 MetArgSerGlyHisGluPheGlyIleGluGlyMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGCCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACCGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAATACCGTAAAGCTTAAAGGTAACCAAGGGGGCCCTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTGTCACCACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGluIleTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGGAGATCTATGAAAGAACCGAGGGACGC 660

Qy  221 HisHisHisLeuPheLeu 225
        |||||||||||||||||
Db  661 CACCATCTGTTCCTT 675
```

RESULT 8
US-10-334-288-6

; Sequence 6, Application US/10334288
; Publication No. US20040002123A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger

APPLICANT: GONZALEZ, Jesus
TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
FILE REFERENCE: REGEN1290-5
CURRENT APPLICATION NUMBER: US/10/334,288
CURRENT FILING DATE: 2002-12-31
PRIOR APPLICATION NUMBER: US/09/967,772
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US 09/459,956
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 08/765,860
PRIOR FILING DATE: 1996-12-19
PRIOR APPLICATION NUMBER: PCT/ US96/09652
PRIOR FILING DATE: 1996-06-06
PRIOR APPLICATION NUMBER: US 08/481,977
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin version 3.0
SEQ ID NO 6
LENGTH: 678
TYPE: DNA
ORGANISM: Discosoma sp "red"
US-10-334-288-6

Alignment Scores:
Pred. No.:
Score: 6.83e-143   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
                            100.00%   Indels: 0
                                            Gaps: 0

US-10-006-922A-12 (1-225) x US-10-334-288-6 (1-678)

```
Qy   1  MetArgSerSerLysGluPheMetArgPheLysValArgMetGluGly          20
Db   1  ATGAGGTCTTCCAAGAAATGTTATCAAGGAGTTTAGAGGTTCGCATGGAAGGA     60

Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db  61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCAGACCATACGAAGGC  120

Qy  41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db 121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db 181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy  81  AspTyrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db 241  GACTATTATAAAAGCTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy 101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db 301  GACGGTGGCGTCGTTACGTTGTATCCTGTGATGCCGGTGTTGCAAGACTTTCATCTAC    360

Qy 121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db 361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGACGGACCTGTTATGCAAAAGAAGACA  420

Qy 141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db 421  ATGGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db 481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  540

Qy 181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db 541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600
```

```
Qy 201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db 601  ATAACAAGCCACCACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGAAGC  660

Qy 221  HisHisLeuPheLeu                                              225
Db 661  CACCATCTGTTCCTT                                              675
```

RESULT 9
US-10-311-030-5
Sequence 5, Application US/10311030
Publication No. US20040171107A1
GENERAL INFORMATION:
APPLICANT: Nelson, David
APPLICANT: Zamlara, Elize
APPLICANT: Tsien, Roger
TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
FILE REFERENCE: 15916-032US1
CURRENT APPLICATION NUMBER: US/10/311,030
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: PCT/US01/04625
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: US 60/184,732
PRIOR FILING DATE: 2000-02-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 678
TYPE: DNA
ORGANISM: Discosoma sp.
US-10-311-030-5

Alignment Scores:
Pred. No.:
Score: 6.83e-143   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
                            100.00%   Indels: 0
                                            Gaps: 0

US-10-006-922A-12 (1-225) x US-10-311-030-5 (1-678)

```
Qy   1  MetArgSerSerLysGluPheMetArgPheLysValArgMetGluGly          20
Db   1  ATGAGGTCTTCCAAGAAATGTTATCAAGGAGTTTAGAGGTTCGCATGGAAGGA     60

Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db  61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCAGACCATACGAAGGC  120

Qy  41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db 121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db 181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy  81  AspTyrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db 241  GACTATTATAAAAGCTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy 101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db 301  GACGGTGGCGTCGTTACGTTGTATCCTGTGATGCCGGTGTTGCAAGACTTTCATCTAC    360

Qy 121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db 361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGACGGACCTGTTATGCAAAAGAAGACA  420

Qy 141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db 421  ATGGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480
```

```
Qy    161  IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db    481  ATTCATGGAGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  540
Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db    541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600
Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db    601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660
Qy    221  HisHisLeuPheLeu  225
Db    661  CACCACCTGTTCCTT  675
```

RESULT 10
US-10-656-029-21
; Sequence 21, Application US/10656029
; Publication No. US20050003367A1
; GENERAL INFORMATION:
; APPLICANT: VERTEX PHARMACEUTICALS INC.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RAPID DEVELOPMENT OF
; TITLE OF INVENTION: SCREENING ASSAYS
; FILE REFERENCE: VPI/02-143WO2
; CURRENT APPLICATION NUMBER: US/10/656,029
; CURRENT FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: 60/408,297
; PRIOR FILING DATE: 2002-09-05
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 21
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(678)
; OTHER INFORMATION: fluorescent protein
US-10-656-029-21

Alignment Scores:
Pred. No.:
Score: 6.83e-143   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 8   Indels: 0
                                           Gaps: 0

US-10-006-922A-12 (1-225) x US-10-656-029-21 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
Db     1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA   60
Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db    61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGATACGAGGC      120
Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIlePro   60
Db   121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db   181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTGAAGCACCCTGCCGACATACCA  240
Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db   241  GACTATAAAAAGCTGTCATTCCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300
Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
```

```
Db   301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGGAGGATGGCTGTTTCATCTAC  360
Qy   121  LysValGlyAsnPheLysIleArgHisAsnValGluAspGlySerValGlnLeuAlaAspHis  140
Db   361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  420
Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db   421  ATGGGCTGGGAAGCCAGCACAGAGCGCCTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  480
Qy   161  IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db   481  ATTCATGGAGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  540
Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db   541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600
Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db   601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660
Qy   221  HisHisLeuPheLeu  225
Db   661  CACCACCTGTTCCTT  675
```

RESULT 11
US-10-505-486-27
; Sequence 27, Application US/10505486
; Publication No. US20050118639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 27
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
US-10-505-486-27

Alignment Scores:
Pred. No.:
Score: 6.83e-143   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 9   Indels: 0
                                           Gaps: 0

US-10-006-922A-12 (1-225) x US-10-505-486-27 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
Db     1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA   60
Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db    61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGATACGAGGC      120
Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIlePro   60
Db   121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
```

RESULT 12

US-10-844-064A-1

```
; Sequence 1, Application US/10844064A
; Publication No. US20050149994A1
; GENERAL INFORMATION:
;  APPLICANT: Bevis, Brooke
;  APPLICANT: Glick, Benjamin
;  TITLE OF INVENTION: RAPIDLY MATURING FLUORESCENT PROTEINS AND METHODS FOR USING THE S
;  FILE REFERENCE: 092234-9006
;  CURRENT APPLICATION NUMBER: US/10/844,064A
;  CURRENT FILING DATE: 2004-05-11
;  PRIOR APPLICATION NUMBER: PCT/US02/40539
;  PRIOR FILING DATE: 2002-12-18
;  PRIOR APPLICATION NUMBER: US 60/341,723
;  PRIOR FILING DATE: 2001-12-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: Patentin version 3.1
;  SEQ ID NO 1
;  LENGTH: 678
;  TYPE: DNA
;  ORGANISM: Discosoma
US-10-844-064A-1
```

US-10-006-922A-12 (1-225) x US-10-844-064A-1 (1-678)

Alignment Scores:

```
Pred. No.:
Score:                 6.838e-143
Score:                 1214.00                Length:        678
Percent Similarity:    100.00%                Matches:       225
Best Local Similarity: 100.00%                Conservative:  0
Query Match:           9                      Mismatches:    0
DB:                                           Indels:        0
                                              Gaps:          0
```

RESULT 13

US-10-931-304-2

```
; Sequence 2, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
;  APPLICANT: Tsien, Roger
;  APPLICANT: Campbell, Robert
;  APPLICANT: Baird, Geoffrey
;  TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;  TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;  FILE REFERENCE: 39754-0831CP2CP3
;  CURRENT APPLICATION NUMBER: US/10/931,304
;  CURRENT FILING DATE: 2004-08-30
;  PRIOR APPLICATION NUMBER: 10/209,208
;  PRIOR FILING DATE: 2002-07-29
;  PRIOR APPLICATION NUMBER: 10/121,258
;  PRIOR FILING DATE: 2002-04-10
;  PRIOR APPLICATION NUMBER: 09/866,538
;  PRIOR FILING DATE: 2001-05-24
;  PRIOR APPLICATION NUMBER: 09/794,308
;  PRIOR FILING DATE: 2001-02-26
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 2
;  LENGTH: 678
;  TYPE: DNA
;  ORGANISM: Discosoma sp.
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(678)
```

; OTHER INFORMATION: wild-type DsRed
US-10-931-304-2

Alignment Scores:
Pred. No.:   Length: 678
Score: 6.83e-143   Matches: 225
Percent Similarity: 1214.00   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 9   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-2 (1-678)

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA  60

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGGGAGGGCCGATACGAAGGC  120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTGTCACCACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTATAAAAAGCTGTCATTCCCTGAAGGATTTAAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA  420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCACTACCTAGTTGAATTCAAAAGTATT  540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT  600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660

Qy  221 HisHisLeuPheLeu  225
        |||||||||||||||
Db  661 CACCATCTGTTCCTT  675
```

US-10-006-922A-12 (1-225) x US-09-999-745-66 (1-859)

Alignment Scores:
Pred. No.:   Length: 859
Score: 9.49e-143   Matches: 225
Percent Similarity: 1214.00   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 3   Gaps: 0

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA  113

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGGGAGGGCCGATACGAAGGC  173

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTCCCTGAAGGATTTAAAATGGGAAAGGGTCATGAACTTTGAA  353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA  473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCACTACCTAGTTGAATTCAAAAGTATT  593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT  653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  713

Qy  221 HisHisLeuPheLeu  225
        |||||||||||||||
Db  714 CACCATCTGTTCCTT  728
```

RESULT 14
US-09-999-745-66

RESULT 15
US-09-866-538-11

```
; Sequence 11, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
;   APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
;   APPLICANT: TSIEN, Roger
;   APPLICANT: Campbell, Robert
;   TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
;   FILE REFERENCE: REGEN1530-2
;   CURRENT APPLICATION NUMBER: US/09/866,538
;   CURRENT FILING DATE: 2001-05-24
;   NUMBER OF SEQ ID NOS: 29
;   SOFTWARE: Patentin version 3.0
; SEQ ID NO 11
;   LENGTH: 859
;   TYPE: DNA
;   ORGANISM: Discosoma sp.
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (54)..(731)
US-09-866-538-11

US-10-006-922A-12 (1-225) x US-09-866-538-11 (1-859)

Alignment Scores:
Pred. No.:
Score:                  1214.00          9.49e-143
Percent Similarity:     100.00%          Length:       859
Best Local Similarity:  100.00%          Matches:      225
Query Match:            100.00%          Conservative:   0
DB:                           3          Mismatches:     0
                                         Indels:         0
                                         Gaps:           0


Qy    1 MetArgSerGlyAlaAsnValIleLysGluMetArgPheLysValArgMetGluGly    20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAAATGTTATCAAGGAGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA  113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly    40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGCGAGGGGCCATACGAAGGC  173

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCCTTGCCATTTGCTTGGGATATT  233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro    80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  713

Qy  221 HisHisLeuPheLeu   225
          |||||||||||||||
Db  714 CACCATCTGTTCCTT  728
```

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 12, 2006, 10:16:32 ; Search time 383 Seconds
                                                     (without alignments)
                                                     475.466 Million cell updates/sec

Copyright (c) 1993 - 2006 Compugen Ltd.
GenCore version 5.1.6

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..........EDYTIVEQYERTEGRHHLPL 225

Scoring table: BLOSUM62
                 Xgapop   10.0 ,   Xgapext   0.5
                 Xgapop   10.0 ,   Xgapext   0.5
                 Fgapop   6.0 ,   Fgapext   0.5
                 Delop   6.0 ,   Delext   7.0

Searched: 6038814 seqs, 404674181 residues

Total number of hits satisfying chosen parameters: 12077628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
Database :   Published Applications NA New:*
1: /cgn2_6/ptcdata/1/pubpna/US08_NEW_PUB.seq:*
2: /cgn2_6/ptcdata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptcdata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptcdata/1/pubpna/US09_NEW_PUB.seq:*
5: /cgn2_6/ptcdata/1/pubpna/US10_NEW_PUB.seq:*
6: /cgn2_6/ptcdata/1/pubpna/US11_NEW_PUB.seq:*
7: /cgn2_6/ptcdata/1/pubpna/US11_NEW_PUB.seq2:*
8: /cgn2_6/ptcdata/1/pubpna/US11_NEW_PUB.seq3:*
9: /cgn2_6/ptcdata/1/pubpna/US11_NEW_PUB.seq4:*
10: /cgn2_6/ptcdata/1/pubpna/US60_NEW_PUB.seq:*

-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/US10006922/runat 10012006 162410 19008/app_query.fasta_1.391
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10006922 @CGN 1 1 184 @runat 10012006 162410 19008
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1214	100.0	678	6	US-10-209-208-2	Sequence 2, Appli
2	1214	100.0	678	7	US-11-218-880-2	Sequence 2, Appli
3	1210	99.7	681	6	US-10-209-208-3	Sequence 3, Appli
4	1210	99.7	681	7	US-10-209-208-23	Sequence 3, Appli
5	1210	99.7	681	7	US-11-218-880-3	Sequence 3, Appli
6	1210	99.7	681	7	US-11-218-880-23	Sequence 23, Appli
7	1186	97.7	6706	6	US-10-655-872-4	Sequence 4, Appli
8	1186	97.7	7927	6	US-10-655-872-7	Sequence 7, Appli

9	1160	95.6	678	6	US-10-209-208-5	Sequence 5, Appli
10	1160	95.6	678	7	US-11-218-880-5	Sequence 5, Appli
11	1121	92.3	681	6	US-10-209-208-7	Sequence 7, Appli
12	1121	92.3	681	7	US-11-218-880-7	Sequence 7, Appli
13	1021	84.1	678	7	US-10-209-208-9	Sequence 9, Appli
14	1012	83.4	678	6	US-10-209-208-80	Sequence 80, Appli
15	486.5	40.1	678	7	US-11-179-411-15	Sequence 15, Appli
16	486.5	40.1	1079	7	US-11-175-766-15	Sequence 15, Appli
17	474.5	39.1	1079	7	US-11-179-411-30	Sequence 30, Appli
18	474.5	39.1	1104	7	US-11-175-766-30	Sequence 30, Appli
19	474.5	39.1	1279	7	US-11-179-411-31	Sequence 31, Appli
20	474.5	39.1	1279	7	US-11-175-766-31	Sequence 31, Appli
21	468.5	38.6	8251	6	US-10-521-768-1	Sequence 1, Appli
22	468.5	38.6	6679	7	US-11-175-766-1	Sequence 1, Appli
23	468.5	38.6	10369	6	US-10-521-768-2	Sequence 2, Appli
24	468.5	38.6	13535	6	US-10-521-768-4	Sequence 4, Appli
25	422.5	34.8	860	7	US-11-179-411-24	Sequence 24, Appli
26	422.5	34.8	860	7	US-11-175-766-24	Sequence 24, Appli
27	422.5	34.8	860	7	US-11-179-411-23	Sequence 23, Appli
28	422.5	34.8	864	7	US-11-175-766-23	Sequence 23, Appli
29	422.5	34.8	864	7	US-11-179-411-26	Sequence 26, Appli
30	422.5	34.8	864	7	US-11-175-766-26	Sequence 26, Appli
31	422.5	34.8	864	7	US-11-179-411-25	Sequence 25, Appli
32	422.5	34.8	873	7	US-11-175-766-25	Sequence 25, Appli
33	422.5	34.8	873	7	US-11-179-411-40	Sequence 25, Appli
34	422.5	34.8	10122	6	US-11-175-766-25	Sequence 25, Appli
35	219.5	18.1	10549	6	US-10-161-408-8	Sequence 8, Appli
36	219.5	18.1	12592	6	US-10-161-408-3	Sequence 3, Appli
37	210.5	17.3	716	6	US-10-209-208-6	Sequence 6, Appli
38	210.5	17.3	716	7	US-10-209-208-21	Sequence 21, Appli
39	210.5	17.3	717	6	US-11-218-880-21	Sequence 21, Appli
40	210.5	17.3	720	7	US-11-089-551A-12	Sequence 12, Appli
41	210.5	17.3	7239	7	US-11-089-551A-39	Sequence 39, Appli
42	209.5	17.3	4773	7	US-11-089-551A-45	Sequence 45, Appli
43	209.5	17.3	5164	7	US-11-214-613-32	Sequence 32, Appli
44	204.5	16.8	1539	7	US-11-214-613-36	Sequence 36, Appli
45	204.5	16.8	3660	7	US-11-032-236-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-209-208-2

; Sequence 2, Application US/10209208
; Publication No. US2005024492IA1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
APPLICANT: Geoffrey Baird
TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
TITLE OF INVENTION: FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP2
CURRENT APPLICATION NUMBER: US/10/209,208
CURRENT FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 80
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 678
TYPE: DNA
ORGANISM: Discosoma sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(678)
OTHER INFORMATION: wild-type DsRed
US-10-209-208-2

Alignment Scores:

Pred. No. :

Score:	2.23e-133	Length: 678
Percent Similarity:	1214.00	Matches: 225
Best Local Similarity:	100.00%	Conservative: 0
Query Match:	100.00%	Mismatches: 0
	6	Indels: 0
DB:		Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-2 (1-678)

```
Qy    1 MetArgSerSerIleAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGCGAGGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValIleLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCACAATTTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspThrTyrLysLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAAAGCTGTCATTTCCTTCCGAAGGATTTAAATGGGAAAGGGTAATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAAGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAACGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAAAAGTCAATAATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

Alignment Scores:

Pred. No. :

Score:	2.23e-133	Length: 678
Percent Similarity:	1214.00	Matches: 225
Best Local Similarity:	100.00%	Conservative: 0
Query Match:	100.00%	Mismatches: 0
	7	Indels: 0
DB:		Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-2 (1-678)

```
Qy    1 MetArgSerSerIleAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGCGAGGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValIleLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCACAATTTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspThrTyrLysLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAAAGCTGTCATTTCCTTCCGAAGGATTTAAATGGGAAAGGGTAATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAAGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAACGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAAAAGTCAATAATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

RESULT 2

RESULT 3

US-10-209-208-3

; Sequence 3, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-10-209-208-3

Alignment Scores:
Pred. No.:
Score: 1210.00
Percent Similarity: 100.00%
Best Local Similarity: 99.56%
Query Match: 6

Length: 681
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-3 (1-681)

Qy   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         ::|:::|:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20

Qy   4 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
         |||||||||||||||
Db 664 CACCACCTGTTCCTG 678

RESULT 4
US-10-209-208-23

; Sequence 23, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-10-209-208-23

Alignment Scores:
Pred. No.:
Score: 1210.00
Percent Similarity: 100.00%
Best Local Similarity: 99.56%
Query Match: 6

Length: 681
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-23 (1-681)

Qy   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         ::|:::|:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20

Qy   4 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Left column

Qy 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 304 GACGCGGGCGGTGGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
Db 664 CACCACCTGTTCCTG 678

RESULT 5
US-11-218-880-3
; Sequence 3, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-11-218-880-3

Alignment Scores:
Pred. No.:
Score: 6.67e-133
Percent Similarity: 1210.00
Best Local Similarity: 100.00%
Query Match: 99.56%
DB: 99.67%

Length: 681
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-3 (1-681)

Qy 1 MetArgSerLeuLysAsnValIleLeuSerGluPheMetArgPheLysValArgMetGluGly 20
Db 4 GTGCGGCTTCCTCCAAGAACGTTCATCAGGCGGTTCAAGGTGCGCCATGGAGGGC 63
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40

Right column

Db 64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTTCGCCCTGCCTGGGACATC 183

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
Db 664 CACCACCTGTTCCTG 678

RESULT 6
US-11-218-880-23
; Sequence 23, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-11-218-880-23

Alignment Scores:

Pred. No.:
Score: 6.67e-133
Percent Similarity: 1210.00
Best Local Similarity: 100.00%
Query Match: 99.56%
DB: 99.67%

Length: 681
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922a-12 (1-225) x US-11-218-880-23 (1-681).

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4 GTGCGCCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCCATGGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheIleLeuIleLysGlyGluGlyGluGlyGluGlyAspAlaThrArgGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGACGCCACCTACGGC 123

Qy   41 HisAsnThrValLeuLysGlyValThrLysGlyLysLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  244 GACACAAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTCCTTCATCTAC 363

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 423

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  664 CACCACCTGTTCCTG 678
```

; FEATURE:
; OTHER INFORMATION: lentiviral vector sequence
US-10-655-872-4
DB:

Alignment Scores:
Pred. No.:
Score: 1e-128
Percent Similarity: 1186.00
Best Local Similarity: 98.22%
Query Match: 97.33%
DB: 97.69%

Length: 6706
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922a-12 (1-225) x US-10-655-872-4 (1-6706)

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2778 ATGCGCTCCTCCAAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCCATGGAGGGC 2837

Qy   21 ThrValAsnGlyHisGluPheIleLeuIleLysGlyGluGlyGluGlyAspAlaThrArgGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2838 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGACGCCACCTACGGC 2897

Qy   41 HisAsnThrValLeuLysGlyValThrLysGlyLysLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2898 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCTTCGCCTGGGACATC 2957

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2958 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3017

Qy   81 AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3018 GACACAAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3077

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3078 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTCCTTCATCTAC 3137

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3138 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 3197

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3198 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3257

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3258 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3317

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||:|||||||||||||||||||||||||||:|||||||||
Db 3318 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 3377

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3378 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 3437

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 3438 CACCACCTGTTCCTG 3452
```

RESULT 7
US-10-655-872-4
; Sequence 4, Application US/10655872
; Publication No. US20050251872A1
; GENERAL INFORMATION:
; APPLICANT: Bear, et al.
; TITLE OF INVENTION: Lentiviral Vectors, Related Reagents, and Methods of Use Thereof
; FILE REFERENCE: 0492611-0512
; CURRENT APPLICATION NUMBER: US/10/655,872
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 4
; LENGTH: 6706
; TYPE: DNA
; ORGANISM: Artificial

RESULT 8
US-10-655-872-7
; Sequence 7, Application US/10655872
; Publication No. US20050251872A1
; GENERAL INFORMATION:
; APPLICANT: Bear, et al.
; TITLE OF INVENTION: Lentiviral Vectors, Related Reagents, and Methods of Use Thereof
; FILE REFERENCE: 0492611-0512
; CURRENT APPLICATION NUMBER: US/10/655,872
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 7
; LENGTH: 7927
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: lentiviral vector sequence
US-10-655-872-7

Alignment Scores:
Pred. No.:   1.26e-128
Score: 1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 97.69%
DB:

Length: 7927
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-655-872-7 (1-7927)

```
Qy    1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
Db 3999  ATGGCCTCTTCCGAGAAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  4058

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db 4059  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  4118

Qy   41  HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db 4119  CACGCCAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  4178

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db 4179  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  4238

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db 4239  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG  4298

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db 4299  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  4358

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db 4359  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC  4418

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db 4419  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  4478

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db 4479  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  4538

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db 4539  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  4598

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db 4599  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  4658

Qy  221  HisHisLeuPheLeu  225
Db 4659  CACCACCTGTTCCTG  4673
```

RESULT 9
US-10-209-208-5
; Sequence 5, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger

; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: US/10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "T1"
US-10-209-208-5

Alignment Scores:
Pred. No.:   5.3e-127
Score: 1160.00
Percent Similarity: 97.33%
Best Local Similarity: 96.00%
Query Match: 95.55%
DB:

Length: 678
Matches: 216
Conservative: 3
Mismatches: 6
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-5 (1-678)

```
Qy    1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
Db    1  ATGGCCTCTTCCGAGGACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db   61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  121  ACCCAGACACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  181  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  241  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  421  ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  600
```

```
Qy  201 IleThrSerHisAsnAspGlnThrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATCACCTCCCACAACGAGGACTACCACCATCGTGGAGCAGTACGAGCGCGACGGAGGGCGGC 660

Qy  221 HisHisLeuPheLeu 225
         |||||||||||||
Db  661 CACCACCTGTTCCTG 675
```

RESULT 10
US-11-218-880-5
; Sequence 5, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: 2002-04-10
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
US-11-218-880-5

Alignment Scores:
Pred. No.:
Score: 5.3e-127   Length: 678
Percent Similarity: 1160.00   Matches: 216
Best Local Similarity: 97.33%   Conservative: 3
Query Match: 96.00%   Mismatches: 6
DB: 95.55%   Indels: 0
                         7   Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-5 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         ||| ||||||||||||||||||: ||| ||||||||||||||||||||
Db   1 ATGGCGTCCTCCGAGGAGAGTCATCAAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
         : ||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||| |||||||||||||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||
Db   ThrValLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
```

```
Qy  201 IleThrSerHisAsnAspGlnThrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATCACCTCCCACAACGAGGACTACCACCATCGTGGAGCAGTACGAGCGCGACGGAGGGCGC 660

Qy  221 HisHisLeuPheLeu 225
         |||||||||||||
Db  661 CACCACCTGTTCCTG 675
```

RESULT 11
US-10-209-208-7
; Sequence 7, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: 10/121,258
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
US-10-209-208-7

Alignment Scores:
Pred. No.:
Score: 2.15e-122   Length: 681
Percent Similarity: 1121.00   Matches: 207
Best Local Similarity: 95.96%   Conservative: 7
Query Match: 92.83%   Mismatches: 9
DB: 92.34%   Indels: 0
                         6   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-7 (1-681)

Qy   3 SerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
         |||||||||||||||||||||||||||||: ||||||||||||||||||:
Db   10 TCCTCCGAGGAGAGTCATCAAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCGTG 69

Qy   23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCAG 129

Qy   43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
         |||
Db 140 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
```

Alignment Scores:

Pred. No.:

Score: Length: 681

Percent Similarity: Matches: 207

Best Local Similarity: Conservative: 7

Query Match: Mismatches: 9

                        Indels: 0

US-11-218-880-7

RESULT 12
US-11-218-880-7
; Sequence 7, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant

2.15e-122

1121.00

95.96%

92.83%

92.34%

DB: 7   Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-7 (1-681)

RESULT 13
US-10-209-208-9
; Sequence 9, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0

Left column

```
; SEQ ID NO 9
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; US-10-209-208-9

Alignment Scores:
Pred. NO.:                              1.36e-110    Length:       678
Score:                                  1021.00      Matches:      192
Percent Similarity:                     89.14%       Conservative: 5
Best Local Similarity:                  86.88%       Mismatches:   24
Query Match:                            84.10%       Indels:       0
DB:                                     6            Gaps:         0

US-10-006-922A-12 (1-225) x US-10-209-208-9 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluGluPheMetArgPheLysValArgMetGluGly  20
        |||
Db   1 ATGGCTTCCTCCGAGAACGTCATCAAGGAGTTCATGGGCTTCAAGGTGCGCATGGAGGGC       60

Qy  21 ThrValAsnGlyHisGluPheGluIleIleGlyGluGlyGluGlyArgProTyrGluGly        40
        ::|
Db  61 GACTACTTGAAGCTGTCCCTCCCGGAGGGCTTCAAGTGGGAGGGCGGGAGGGCCTACGAGGGC    120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle        60
        |||
Db 121 GACGGCGGCGGCGTGGTGACCGTGACCAAGGGCGGCCCCCCCTTGCCTTGGGACATC         180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro       80
        ||
Db 181 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC      240

Qy  81 AspThrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu          100
        ||
Db 241 GACTACTTGAAGCTGTCCTTCCCCGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG     300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr       120
        |||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC      360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr       140
        |||
Db 361 AAGGTGAAGCTGCGCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC      420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu       160
        |||
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGAGACGGCGTGCTGAAGGGCGAG      480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle       180
        |||:
Db 481 ATCAAGATGAGGGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCACC     540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp       200
        |||:::
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC      600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg       220
        |||
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGC      660

Qy 221 His    221
        |||
Db 661 CAC    663
```

Right column

APPLICANT: Campbell, Robert

```
; SEQ ID NO 9
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; US-11-218-880-9

Alignment Scores:
Pred. NO.:                              1.36e-110    Length:       678
Score:                                  1021.00      Matches:      192
Percent Similarity:                     89.14%       Conservative: 5
Best Local Similarity:                  86.88%       Mismatches:   24
Query Match:                            84.10%       Indels:       0
DB:                                     7            Gaps:         0

US-10-006-922A-12 (1-225) x US-11-218-880-9 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluGluPheMetArgPheLysValArgMetGluGly  20
        |||
Db   1 ATGGCTTCCTCCGAGAACGTCATCAAGGAGTTCATGGGCTTCAAGGTGCGCATGGAGGGC       60

Qy  21 ThrValAsnGlyHisGluPheGluIleIleGlyGluGlyGluGlyArgProTyrGluGly        40
        ::|
Db  61 TCCGGTGAACGGCCACGAGTTCGAGATCGAGGGCGGCGAGGGCGCGAGGGCCTACGAGGGC     120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle        60
        |||
Db 121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCCCTTGCCTTGGGACATC         180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro       80
        ||
Db 181 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC      240

Qy  81 AspThrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu          100
        ||
Db 241 GACTACTTGAAGCTGTCCTTCCCCGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG     300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr       120
        |||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC      360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr       140
        |||
Db 361 AAGGTGAAGCTGCGCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC      420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu       160
        |||
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGAGACGGCGTGCTGAAGGGCGAG      480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle       180
        |||:
Db 481 ATCAAGATGAGGGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCACC     540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp       200
        |||:::
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC      600
```
```
RESULT 14
US-11-218-880-9
; Sequence 9, Application US/11218880
; Publication No US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Talen, Roger
```

```
Qy    201 IleThrSerHisAsnGluAspThrTyrThrIleValGlnGlnTyrGluGluGlyAsp  220
          ATCACCTCCCACAACGAGGACTACACCATCGTGCAACAGTACGAGGAGGGCGAGGGCCGC  660
Db    601 |||                                                           663
          CAC                                                           663
          His  221

Qy    221 His  221
Db    661 CAC  663
          His  221

RESULT 15
US-10-209-208-80

; Sequence 80, Application US/10209208
; Publication No. US20050249215A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR APPLICATION FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR APPLICATION FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR APPLICATION FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 4.0
; SEQ ID NO 80
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "mRFP1.1"
US-10-209-208-80

Alignment Scores:
Pred. No.:         1.58e-109
Score:             1012.00       Length:        678
Percent Similarity: 88.69%       Matches:       190
Best Local Similarity: 85.97%    Conservative:  6
Query Match:        83.36%       Mismatches:    25
DB:                6             Indels:        0
                                 Gaps:          0

US-10-006-922A-12 (1-225) x US-10-209-208-80 (1-678)

Qy      1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly  20
          |||||||||||| || ||||||||||| |||||||||||||| ||||||||
Db      1 ATGGCCTCTCCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
          ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
          ||||                |||||||||||||||||||||||||||||
Db    121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
          |||||||| |||||  ||||||||||||||||||||||||||||||
Db    181 CTGTCCCCTCAGTTCCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC  240

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2006 Compugen Ltd.
GenCore version 5.1.6

Run on: January 11, 2006, 02:04:28 ; Search time 77 Seconds
(without alignments)
1283.899 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRPKVRMEG........EDYIVEQVERTEGRHHLFL 225

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 43978781 residues

Total number of hits satisfying chosen parameters: 109

Post-processing: Minimum Match 69%
                 Maximum Match 100%
                 Listing first 500 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :
       A_Geneseq_21:*
       1: geneseqp1980s:*
       2: geneseqp1990s:*
       3: geneseqp2000s:*
       4: geneseqp2001s:*
       5: geneseqp2002s:*
       6: geneseqp2003s:*
       7: geneseqp2003bs:*
       8: geneseqp2004s:*
       9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	225	3	AAY99836	Discosoma
2	1214	100.0	225	7	AAB01622	Discosoma
3	1214	100.0	225	4	AAG65509	Anthozoan
4	1214	100.0	225	5	ABB08834	Yeast opt
5	1214	100.0	225	5	AAE28833	Discosoma
6	1214	100.0	225	7	AAE17540	Discosoma
7	1214	100.0	225	5	AAO18270	Discosoma
8	1214	100.0	225	6	ADY51716	Discosoma
9	1214	100.0	225	7	AAE34962	Discosoma
10	1214	100.0	225	7	ADC24126	Discosoma
11	1214	100.0	225	7	ABW00918	Discosoma
12	1214	100.0	225	7	ADF70403	Discosoma
13	1214	100.0	225	8	ADH34489	Discosoma
14	1214	100.0	225	7	ADL46203	Discosoma
15	1214	100.0	225	8	ADN33979	Wild-type
16	1214	100.0	225	8	ADI36421	Discosoma
17	1214	100.0	225	8	ADM97769	D sp red
18	1214	100.0	225	9	ADQ28780	Jellyfish
19	1214	100.0	225	5	ADX26534	Discosoma
20	1214	100.0	487	6	ABB08821	Autofluor
21	1214	100.0	506	5	ABB08822	Autofluor
22	1214	100.0	547	5	ABB08823	Autofluor
23	1214	99.8	225	5	AAE28920	Discosoma
24	1211	99.8	225	7	ABW00937	Discosoma
25	1211	99.8	225	7	ABW00929	Discosoma
26	1211	99.8	225	7	ABW00938	Discosoma
27	1211	99.8	225	5	ADH34498	Discosoma
28	1211	99.8	226	5	ADH34499	Discosoma
29	1211	99.8	225	5	AAE28919	Discosoma
30	1210	99.7	225	7	ABW00930	Discosoma
31	1210	99.7	226	4	AAG65510	Anthozoan
32	1210	99.7	242	7	ADE24109	Discosoma
33	1210	99.7	545	7	ADL18132	RFP:PS (NI
34	1210	99.7	548	5	ADL18156	RFP:PS (HI
35	1209	99.6	225	7	AAE28922	Discosoma
36	1209	99.6	225	5	ABW00936	Discosoma
37	1209	99.6	225	5	ABW00931	Discosoma
38	1208	99.6	226	5	ABB08835	Yeast opt
39	1208	99.6	225	7	ABW00932	Discosoma
40	1208	99.5	225	7	ABW00935	Discosoma
41	1208	99.5	225	7	ABW00939	Discosoma
42	1207	99.4	225	7	ADH34500	Discosoma
43	1207	99.4	225	5	AAE28921	Discosoma
44	1207	99.4	225	7	AAE17541	Discosoma
45	1207	99.4	225	7	ADY51734	Discosoma
46	1207	99.4	225	7	ABW00941	Discosoma
47	1207	99.4	225	5	ABW00940	Discosoma
48	1206	99.4	225	7	ADL46222	Discosoma
49	1206	99.3	225	5	ABW00933	Discosoma
50	1205	99.3	225	7	ADH34501	Discosoma
51	1205	99.3	240	6	ABP56678	Mammalian
52	1205	99.3	236	9	AAE28923	Discosoma
53	1204	99.2	225	5	AAE28925	Discosoma
54	1204	99.2	225	7	AAE28924	Discosoma
55	1202	99.0	225	7	ADH34503	Discosoma
56	1202	99.0	225	7	ADH34502	Discosoma
57	1202	99.0	240	6	ABP56685	Discosoma
58	1200	99.0	240	6	ABP70039	Colour Fa
59	1200	98.8	225	5	AAE28926	Discosoma
60	1199	98.8	225	7	ADH34505	Discosoma
61	1199	98.8	225	7	ADH34505	Discosoma
62	1199	98.8	225	7	ABP56684	Discosoma
63	1199	98.8	240	6	ABP56683	Discosoma
64	1199	98.5	225	7	ADH34504	Discosoma
65	1196	98.5	236	8	ADO78068	Corallimo
66	1196	98.5	240	6	ADO78070	Corallimo
67	1196	98.5	236	9	ADZ84223	Actinodis
68	1196	98.5	236	9	ADZ84225	Actinodis
69	1195.5	98.4	284	8	ADQ59962	Discosoma
70	1194	98.4	240	6	ABP56681	Discosoma
71	1194	98.3	225	6	ADH34506	Discosoma
72	1193	98.3	240	8	ABP56679	Discosoma
73	1192	98.2	240	6	ABP56680	Discosoma
74	1191	98.2	225	7	AAE28837	Discosoma
75	1191	98.1	225	5	AAE17542	Discosoma
76	1188	97.9	240	6	ABP56682	Discosoma
77	1186	97.7	225	8	AAE28836	Discosoma
78	1186	97.7	225	7	ADH10227	Red fluor
79	1186	97.7	225	8	ADT77526	Protein S
80	1186	97.7	239	8	ADH10228	Immobilis
81	1186	97.7	239	8	ADT77528	Protein/1
82	1183	97.4	240	6	ABP56686	Discosoma
83	1175.5	96.8	226	8	ADL26532	Discosoma
84	1160	95.6	225	7	ADC24129	Discosoma
85	1160	95.6	225	7	ADL46206	Discosoma
86	1153	95.0	225	8	ADM13537	Superfold
87	1132	93.2	226	7	ADL46208	Discosoma
88	1121	92.3	226	7	ADC24131	Discosoma
89	1121	92.3	226	3	ADL46208	Red fluor
90	1119	92.2	225	5	AAY99838	Humanised
91	1119	92.2	225	5	AAE28835	Discosoma
92	1111	91.5	225	3	AAY99839	Humanised
93	1111	91.5	223	5	AAE28828	Discosoma
94	1106	91.1	223	5	AAE28929	Discosoma
95	1105	91.0	225	5	AAY99840	Humanised
96	1105	91.0	225	5	AAE28927	Discosoma
97	1085.5	89.4	230	3	AAY92894	Discosoma

98	1085.5	89.4	230	3	Aab01625 Discosoma
99	1085.5	89.4	230	4	Aae00378 Discosoma
100	1085.5	89.4	230	5	Abp70038 Colour Fa
101	1085.5	89.4	230	6	Abp96651 Red fluor
102	1058	87.1	230	6	Aae00375 Discosoma
103	1054	86.8	205	4	Aae00384 Discosoma
104	1053	86.7	205	4	Aae00383 Discosoma
105	1021	84.1	225	7	Adc24133 Discosoma
105	1021	84.1	225	7	Adl46210 Discosoma
106	1021	84.1	225	7	Adq59566 Discosoma
107	1021	84.1	225	8	Aea54910 Fluoresce
108	1021	84.1	747	8	Adl46281 Discosoma
109	1012	83.4	225	7	

ALIGNMENTS

RESULT 1
AAY99836
ID   AAY99836 standard; protein; 225 AA.
XX
AC   AAY99836;
XX
DT   12-SEP-2003 (revised)
DT   19-SEP-2000 (first entry)
XX
DE   Discosoma sp. "red" novel fluorescent protein drFP583.
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling.
XX
OS   Discosoma sp. "red".
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 122
FT   /notes= "encoded by TC"
FT   Misc-difference 127
FT   /notes= "encoded by GTTG"
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029473.
XX
PR   11-DEC-1998; 98US-00210330.
PR   14-OCT-1999; 99US-0041B529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
DR   WPI; 2000-423381/36.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 20; Page 74-75; 86pp; English.
XX
CC   The present sequence is a novel fluorescent protein (nFP) encoded by the
CC   full-length cDNA drFP583. drFP583 was isolated from Discosoma sp. "red",
CC   a non-bioluminescent species of the Class Anthozoa. Fluorescent proteins
CC   can be used in fluorescent labeling, a useful tool for marking a protein,
CC   cell or organism of interest. Unlike other markers used in protein
CC   labeling, such as beta-galactosidase and luciferase, fluorescent proteins
CC   do not require an exogenous cofactor or substrate. Methods involving
CC   fluorescent proteins are also less laborious and less difficult to
CC   control than the traditional methods of fluorescent labeling, where a
CC   protein of interest is purified and then covalently conjugated to a
CC   fluorophore derivative. Novel fluorescent proteins isolated from species
CC   of the Class Anthozoa can be used as markers for gene expression and
CC   protein localization studies, and in fluorescence resonance energy
CC   transfer (FRET) reactions. They may have improved properties and better
CC   suitability for larger excitations compared to prior art fluorescent
CC   proteins such as green fluorescent protein. (Updated on 12-SEP-2003 to
CC   standardise OS field)
XX
SQ   Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 3; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 2
AAB01622
ID   AAB01622 standard; protein; 225 AA.
XX
AC   AAB01622;
XX
DT   12-DEC-2000 (first entry)
XX
DE   Discosoma sp. red fluorescent protein drFP583.
XX
KW   Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;
KW   zFP506; zFP538; drFP583; dsFP483; asFP600; dgFP512; dmFP592.
XX
OS   Discosoma sp.
XX
PN   WO200034526-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029405.
XX
PR   11-DEC-1998; 98US-00210330.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
XX
DR   WPI; 2000-423451/36.
XX
PT   Novel method for identifying a DNA sequence encoding fluorescent proteins
PT   from non-bioluminescent Anthozoa which are useful for fluorescent
PT   labeling and as markers.
XX
PS   Claim 3; Page 68-69; 73pp; English.
XX
CC   The present sequence is Discosoma sp. red fluorescent protein drFP583. It
CC   was isolated using the Aequoria victoria green fluorescent protein (GFP)
CC   sequence, which was used to design PCR primers which might isolate other
CC   fluorescent proteins from a number of species of Anthozoa. These were
CC   Anemonia majano, Clavularia sp. Zoanthus sp. Discosoma sp. and Anemonia
CC   sulcata. The cDNA obtained was then screened in the search for sequences
CC   encoding fluorescent proteins. The other proteins found in this manner
CC   were cFP484, zFP506, zFP538, amFP486, dsFP483, asFP600, dgFP512 and
CC   dmFP592. These proteins can be used as fluorescence labels (for gene
CC   expression and protein localisation studies and in fluorescence resonance

Qy   61 LSPQFQYGSKVVVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||| ||||||||||||||| |||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

CC energy transfer (FRET) studies) in place of fluorophore derivatives and
CC luciferases, as these involve laborious processes and the latter require
CC cofactors. They can also be used in place of GFP, which is too stable to
CC be useful when studying short-term or repetitive events
XX
SQ Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 3; Length 225;
Best Local Similarity   100.0%; Pred. No. 1e-127;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

(truncated sequence listing content)

```
CC   sequence is that of the yeast optimised RFP
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 5;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;            Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 5

```
ID   AAE28833
AC   AAE28833;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) wild-type protein.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescence timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6.
XX
OS   Discosoma sp.
XX
PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-0006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savietsky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   N-PSDB; AAD46278.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page 70-71; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. drFP583 (NFP-6) wild-type protein of the
CC   invention
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 5;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;            Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 6

```
ID   AAE17540
AC   AAE17540;
XX
DT   22-APR-2002  (first entry)
XX
DE   Discosoma sp. humanised wild-type Anthozoa protein drFP583.
XX
KW   Fluorescent timer protein; protein movement; translocation; trafficking;
KW   promoter activity; gene expression; transgenic plant; gene modification;
KW   protein age; anthozoa protein; drFP583.
XX
OS   Discosoma sp.
XX
PN   WO200196373-A2.
XX
PD   20-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-US019097.
PR   14-JUN-2000; 2000US-0211607P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Fradkov AF, Terskikh A;
XX
DR   WPI; 2002-154595/20.
DR   N-PSDB; AAD28207.
XX
PT   New fluorescent timer proteins comprising an emission spectrum that
PT   changes over time from a first wavelength to a second wavelength, useful
PT   for monitoring intracellular protein movement, translocation, trafficking
PT   or stability.
XX
PS   Example 1; Fig 1; 89pp; English.
XX
CC   The invention relates to a fluorescent timer protein having an emission
CC   spectrum that changes over time after synthesis from a first wavelength
CC   to a second wavelength. The fluorescent timer proteins are useful in
CC   monitoring the activity of a promoter, determining the age of a protein,
CC   identifying an agent that modulates the activity of a promoter and in
CC   enriching a population of cells comprising a fluorescent timer protein.
CC   The fluorescent timer proteins are also useful for assessing gene
CC   expression during development of a multicellular organism or during
```

CC cellular differentiation, in response to a drug or other inducer of
CC promoter activity, as a reporter to serve as a read-out of promoter
CC activity, monitoring intracellular protein movement or translocation,
CC protein trafficking, or protein stability, to investigate temporal
CC aspects of the activity of a regulatory element, for determining cell
CC fate during development and organ remodelling, in spatial and temporal
CC visualisation of newly synthesised proteins and accumulated proteins, and
CC in distinguishing between newly formed and pre-existing structures, e.g.
CC membrane junctions and extracellular matrix components. The fluorescent
CC timer proteins may further be used to investigations where photobleaching
CC techniques are employed, as detectable labels, as selectable markers, as
CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC assays, and as second messenger detectors. The nucleic acids can be used
CC to generate transgenic, non-human plants or animals or site-specific gene
CC modifications in cell lines. The present sequence is Discosoma sp.
CC humanised wild-type Anthozoa protein drFP583 used for generating
CC fluorescent proteins
XX
SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 7
AAO18270
ID AAO18270 standard; protein; 225 AA.
XX
AC AAO18270;
XX
DT 26-SEP-2002 (first entry)
XX
DE Discosoma red fluorescent protein.
XX
KW Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
KW modified yeast strain; environmental pollution.
XX
OS Discosoma sp.
XX
PN DE10061872-A1.
XX
PD 20-JUN-2002.
XX
PF 12-DEC-2000; 2000DE-01061872.
PR 12-DEC-2000; 2000DE-01061872.
XX
PA (LICH/) LICHTENBERG-FRATE H.
XX
PI Lichtenberg-Frate H;
XX
DR WPI; 2002-539633/58.
DR N-PSDB; AAL47952.
XX
PT Modified yeast strain, useful for detecting toxic compounds in
PT environment, contains integrated cassettes responsive to genotoxic and

---

cytotoxic compounds.
XX
PS Disclosure; Page 21-22; 34pp; German.
XX
CC The present invention relates to a modified yeast strain that contains,
CC integrated stably and functionally in its genome, a genotoxicity cassette
CC and a cytotoxicity cassette, each comprising a promoter and reporter
CC gene, both of which are different in the two cassettes. The modified
CC yeast strain is used to detect environmental pollution, especially
CC genotoxic and/or cytotoxic substances in complex environmental
CC contaminants, especially organic compounds, but also (non-)ionising
CC radiation and chemical carcinogens. Particular applications are in
CC monitoring (waste) water (e.g. as an early warning system), medical
CC toxicology screening and for industrial process control. The present
CC sequence is a marker protein suitable for use in the cassettes of the
CC present invention
XX
SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e-127;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 8
ADY51716
ID ADY51716 standard; protein; 225 AA.
XX
AC ADY51716;
XX
DT 05-MAY-2005 (first entry)
XX
DE Discosoma sp. wild type red fluorescent protein (RFP) Seq 12.
XX
KW fluorescence; mutagenesis; red fluorescent protein; protein interaction.
XX
OS Discosoma sp.
XX
PN WO200268605-A2.
XX
PD 06-SEP-2002.
XX
PF 26-FEB-2002; 2002WO-US006063.
XX
PR 26-FEB-2001; 2001US-00794308.
PR 24-MAY-2001; 2001US-00866538.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Baird GS, Campbell RE, Zacharias DA;
XX
DR WPI; 2002-713372/77.
DR N-PSDB; ADY51715.
XX
PT New non-oligomerizing fluorescent protein containing at least one
PT mutation that reduces or eliminates the ability of the protein to
PT oligomerize, useful for making better and new assays for molecular

biology.
XX
PS Claim 10; SEQ ID NO 12; 117pp; English.
XX
PT This invention relates to a novel non-oligomerizing fluorescent protein.
CC Specifically, it refers to the presence of at least one mutation in the
CC fluorescent protein that reduces or eliminates the ability of the protein
CC to oligomerize. The present invention describes fluorescent proteins
CC derived from naturally occurring green or red fluorescent proteins and
CC provides a fusion protein that comprises a non-oligomerizing fluorescent
CC protein linked to at least one protein of interest. As such, these fusion
CC proteins can be used in methods and compositions to determine the pH of a
CC sample, or whether the sample contains an enzyme, molecule or agent that
CC regulates the activity of an expression control sequence. Furthermore,
CC they may be used to identify a specific interaction of molecules, such
CC that they are useful for improving or developing new assays in the field
CC of molecular biology. This polypeptide sequence is the Discosoma sp. wild
CC type red fluorescent protein (RFP) of the invention.
XX
SQ Sequence 225 AA;
XX

```
Query Match             100.0%;  Score 1214; DB 5; Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 9
AAE34962
ID AAE34962 standard; protein; 225 AA.
XX
AC AAE34962;
XX
DT 28-MAY-2003 (first entry)
XX
DE Discosoma species red fluorescent protein (RFP).
XX
KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
KW kinase; red fluorescent protein; RFP.
XX
OS Discosoma sp.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PN WO20029058-A2.
XX
PD 28-NOV-2002.
XX
PF 24-MAY-2002; 2002WO-US016955.
XX
PR 24-MAY-2001; 2001US-00865291.
XX
PI Tsien RY, Ting AY, Zhang J;
XX
DR WPI; 2003-148474/14.
DR N-PSDB; AAD53432.
XX
PT Novel chimeric phosphorylation indicators, useful for detecting

PT kinase/phosphatase in samples, has donor molecule, phosphorylatable
PT domain, phosphoaminoacid binding domain, and acceptor molecule, in
PT operative linkage.
XX
PS Disclosure; Col 65-66; 38pp; English.
XX
PT The present invention relates to chimeric phosphorylation indicators
CC comprising a phosphorylation polypeptide and a fluorescent protein or
CC operative linkage, a phosphorylatable domain, a
CC phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The
CC phosphorylation indicators of the invention are useful for detecting
CC kinases or phosphatases in a biological sample. They are also useful in
CC high throughput analysis e.g. for detecting a kinase inhibitor or
CC phosphatase inhibitor. The present sequence is Discosoma species red
CC fluorescent protein (RFP) used in the invention
XX
SQ Sequence 225 AA;
XX

```
Query Match             100.0%;  Score 1214; DB 6; Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 10
ADC24126
ID ADC24126 standard; protein; 225 AA.
XX
AC ADC24126;
XX
DT 18-DEC-2003 (first entry)
XX
DE Discosoma wild-type red fluorescent protein.
XX
KW Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW fluorescent protein variant; transcription induction detection;
KW fluorescence energy resonance transfer; FRET; protein kinase;
KW protein phosphatase; ion indicator.
XX
OS Discosoma.
XX
PN US2003059835-A1.
XX
PD 27-MAR-2003.
XX
PF 10-APR-2002; 2002US-00121258.
XX
PR 26-FEB-2001; 2001US-00794308.
PR 24-MAY-2001; 2001US-00865538.
XX
PA (TSIE/) TSIEN R Y.
PA (CAMP/) CAMPBELL R E.
XX
PI Tsien RY, Campbell RE;
XX
DR WPI; 2003-743764/70.
DR N-PSDB; ADC24127, ADC24134.
XX

PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.
XX
PS   Claim 1; SEQ ID NO 1; 67pp; English.
XX
CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   containing a vector which comprises (I) operatively linked to an
CC   expression control sequence, and an unit to assay the variant
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescent protein produced by (VIII), where variant fluorescent protein
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcriptions, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This is the amino acid sequence of
CC   Discosoma wild-type red fluorescent protein.
XX
SQ   Sequence 225 AA;
XX

Query Match   100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity   100.0%;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 11
ID   ABW00918
AC   ABW00918;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP).
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP.
OS   Discosoma sp.

XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
DR   N-PSDB; AAD61969.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Claim 10; Page 30-31; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP)
XX
SQ   Sequence 225 AA;
XX

Query Match   100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity   100.0%;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 12
ID   ADF70403
AC   ADF70403;
XX
DT   12-FEB-2004 (first entry)
XX
DE   Discosoma wild-type GFP variant protein SeqID26.
XX
KW   ligand; orphan receptor protein; fusion protein; fluorescent protein;
KW   cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
KW   GFPuv; Enhanced GFP; EGFP.
OS   Discosoma sp.
XX
PN   WO200307l272-A1.

XX
PD   28-AUG-2003.
XX
PF   21-FEB-2003; 2003WO-JP001901.
XX
PR   22-FEB-2002; 2002JP-00045728.
PR   23-JUL-2002; 2002JP-00213949.
PR   11-OCT-2002; 2002JP-00298237.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX
DR   WPI; 2003-697654/66.
DR   N-PSDB; ADF70404.
XX
PT   Transformation of cells with a fusion protein of an orphan receptor
PT   protein with a fluorescent protein useful for identification of ligands
PT   to the orphan receptor.
XX
PS   Disclosure; SEQ ID NO 26; 594pp; Japanese.
XX
CC   This invention relates to a novel method of identifying ligands to an
CC   orphan receptor protein which comprises transforming cells with DNA
CC   encoding a fusion protein of the orphan receptor with a fluorescent
CC   protein, so that the fusion protein is expressed in the cells (or cell
CC   membranes isolated from them) and contacting the cells with the potential
CC   ligand to be tested. A suitable fluorescent protein for incorporation in
CC   the fusion protein is green fluorescent protein (GFP), for example GFP-1,
CC   wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
CC   identification of ligands binding to an orphan receptor protein.
XX
SQ   Sequence 225 AA;

Query Match   100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity   100.0%; Pred. No. 1e-127;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKYTVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKYTVKHPADIPDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 13

ID   ADH34489   standard; protein; 225 AA.
XX
AC   ADH34489;
XX
DT   11-MAR-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein DsRed (wild-type).
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agent; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   fluorescence activated cell sorting; fluorescence timer;
KW   red fluorescent protein; DsRed.
XX

OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 146
FT   /note= "Key position for chromoprotein/fluorescent
FT   protein activity. Corresponds to GFP residue 148"
FT   Misc-difference 161
FT   /note= "Key position for chromoprotein/fluorescent
FT   protein activity. Corresponds to GFP residue 165"
FT   Misc-difference 163
FT   /note= "Key position for chromoprotein/fluorescent
FT   protein activity. Corresponds to GFP residue 167"
FT   Misc-difference 197
FT   /note= "Key position for chromoprotein/fluorescent
FT   protein activity. Corresponds to GFP residue 203"
XX
PN   WO2003057833-A2.
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.
XX
PR   26-DEC-2001; 2001US-0343128P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo- or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Fig 1; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably a non-bioluminescent
CC   Cnidarian species, and most preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   the invention. The interconverted mutant proteins of the invention are
CC   useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can be useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as sunscreens and selective filters. Chromoprotein
CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g., transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g., green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence

CC   represents a wild-type red fluorescent protein, DsRed, from Discosoma sp.
CC   that was used as a parent sequence for the generation of mutant proteins
CC   in an example of the invention.
XX
SQ   Sequence 225 AA;

```
Query Match             100.0%;  Score 1214;  DB 7;  Length 225;
Best Local Similarity   100.0%;
Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 14
ADL46203
ID   ADL46203 standard; protein; 225 AA.
XX
AC   ADL46203;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Discosoma red fluorescent (DsRed) protein.
XX
KW   red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
XX
PN   WO200308646446-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010879.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-00209208.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Tsien RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
DR   N-PSDB; ADL46204.
XX
PS   Claim 1; SEQ ID NO 1; 166pp; English.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitution at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced

CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology, and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   protein.
XX
SQ   Sequence 225 AA;

```
Query Match             100.0%;  Score 1214;  DB 7;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 15
ADN33979
ID   ADN33979 standard; protein; 225 AA.
XX
AC   ADN33979;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Wild-type DsRED protein.
XX
KW   Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED.
XX
OS   Discosoma sp.
XX
XX
PN   WO200305415B-A2.
XX
PD   03-JUL-2003.
XX
PF   18-DEC-2002; 2002WO-US040539.
XX
PR   19-DEC-2001; 2001US-0341723P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Bevis B, Glick B;
XX
DR   WPI; 2003-569236/53.
DR   N-PSDB; ADN33978.
XX
PS   Claim 8; SEQ ID NO 2; 65pp; English.
XX
PT   Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT   mutant of a Cnidarian chromo- or fluorescent protein or its mutant,
PT   useful for applications involving chromo- or fluorescent proteins.
XX
CC   The present invention relates to nucleic acid that encodes a rapidly
CC   maturing chromo or fluorescent mutant of a Cnidarian chromo- or
CC   fluorescent protein or its mutant. The protein is useful in applications
CC   involving nucleic acid encoding a chromo- or fluorescent protein and is
CC   useful for producing a chromo and/or fluorescent protein which involves
CC   growing the cell, whereby the protein is expressed, and isolating the
CC   protein substantially free of other proteins. The protein is useful in
CC   applications involving chromo- or fluorescent protein and is useful as
CC   PCR primers, hybridization probes, etc. The expression cassettes are

CC   useful for synthesizing related proteins. The chromoproteins are useful
CC   as coloring agents which are capable of imparting color or pigment to a
CC   particular composition of matter e.g. food compositions, pharmaceuticals,
CC   cosmetics, living organisms, e.g., animals and plants. The chromoproteins
CC   may also find use as labels in analyte detection assays, e.g. assays for
CC   biological analytes of interest and as selectable markers in recombinant
CC   DNA applications, e.g. the production of transgenic cells and organisms.
CC   The fluorescent proteins find use in a variety of different applications,
CC   e.g. in fluorescence resonance energy transfer (FRET) applications, as
CC   biosensors in prokaryotic and eukaryotic cells, in applications involving
CC   the automated screening of arrays of cells expressing fluorescent
CC   reporting groups by using microscopic imaging and electronic analysis, as
CC   second messenger detectors, and in fluorescence activated cell sorting
CC   applications and as in vivo marker in animals. The fluorescent proteins
CC   also find use in protease cleavage assays. The proteins can also be used
CC   is assays to determine the phospholipid composition in biological
CC   membranes and as a fluorescent timer. The present sequence represents the
CC   wild-type DsRED.
XX
SQ   Sequence 225 AA;

Query Match   Best Local Similarity   100.0%;
                        Score 1214; DB 7; Length 225;
Matches 225; Conservative   0; Pred. No. 1e-127; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 16
AD136421
ID   AD136421 standard; protein; 225 AA.
XX
AC   AD136421;
XX
DT   22-APR-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RED).
XX
KW   Fluorobody; binding ligand; green fluorescent protein; GFP;
KW   target detection; red fluorescent protein; RED.
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FT   Region   22..26
FT   /note= "Loop region"
FT   Region   37..39
FT   /note= "Loop region"
FT   Region   75..81
FT   /note= "Loop region"
FT   Region   100..103
FT   /note= "Loop region"
FT   Region   114..117
FT   /note= "Loop region"
FT   Region   153..156
FT   /note= "Loop region"
FT   Region   167..170
FT   /note= "Loop region"
FT   Region   185..192
FT   /note= "Loop region"

FT   Region   204..209
FT   /note= "Loop region"
XX
PN   US20030203355-A1.
XX
PD   30-OCT-2003.
XX
PF   24-APR-2002; 2002US-00132067.
XX
PR   24-APR-2002; 2002US-00132067.
XX
PA   (LALA-) LOS ALAMOS NAT LAB.
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Bradbury AM, Zeytun A, Waldo GS;
XX
DR   WPI; 2004-154325/15.
DR   N-PSDB; AD136420.
XX
PS   Claim 2; SEQ ID NO 4; 23pp; English.
XX
CC   The invention relates to binding ligands (fluorobodies) with intrinsic
CC   fluorescence, which comprises green fluorescent protein (GFP) having
CC   heterologous binding sites. The binding ligand is useful for detecting
CC   the target molecule and is efficiently detects the target molecule. The
CC   present sequence is Discosoma sp. red fluorescent protein (RED) used in
CC   the exemplification of the invention.
XX
SQ   Sequence 225 AA;

Novel binding ligand with intrinsic fluorescence and comprising
fluorescent protein having heterologous binding sites, useful for
detecting target molecule.

Query Match   Best Local Similarity   100.0%;
                        Score 1214; DB 8; Length 225;
Matches 225; Conservative   0; Pred. No. 1e-127; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 17
ADM97769
ID   ADM97769 standard; protein; 225 AA.
XX
AC   ADM97769;
XX
DT   01-JUL-2004 (first entry)
XX
DE   D sp red fluorescent protein SEQ ID NO: 22.
XX
KW   enzyme; sensor cell; green fluorescent protein;
KW   signal transduction detection system; promoter; targeting sequence;
KW   targeted drug.
OS   Discosoma sp.
XX
PN   WO2004031415-A2.

RESULT 18

```
RESULT 18
ADQ28780
ID   ADQ28780 standard; protein; 225 AA.
AC   ADQ28780;
XX
DT   07-OCT-2004   (first entry)
XX
DE   Jellyfish green fluorescent protein DsRed.
XX
KW   Green fluorescent protein; fluorescent protein.
OS   Aequorea victoria.
XX
PN   WO200405973-A1.
XX
PD   15-JUL-2004.
XX
PD   15-APR-2004.
XX
PF   05-SEP-2003; 2003WO-US028078.
XX
PR   05-SEP-2002; 2002US-0408297P.
XX
PA   (VERT-) VERTEX PHARM INC.
XX
PI   Whitney MA, Zeh K, Sanders PS;
XX
DR   WPI; 2004-330208/30.
DR   N-PSDB; ADM97768.
XX
PT   Developing a sensor cell, useful in determining the activity of a target
PT   gene and in developing therapeutic drugs, comprises providing cells
PT   comprising a signal transduction detection system and introducing DNA
PT   construct into cells.
XX
PS   Disclosure; Page 168-169; 234pp; English.
XX
CC   The present invention relates to a method of developing a sensor cell,
CC   for determining the activity of a target gene in the cell, which
CC   comprises providing a homogeneous population of cells, where each of the
CC   cells comprises a signal transduction detection system and introducing
CC   into the population of cells an isolated DNA construct comprising a
CC   promoter operatively linked to a targeting sequence. The method is useful
CC   in developing a sensor cell for determining the activity of a target gene
CC   in the cell. The sensor cell and the methods are useful in developing new
CC   and therapeutic drugs directed to the targets. The present sequence is a
CC   polypeptide shown in the exemplification of the invention.
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 8;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 19

```
RESULT 19
ADX26534
ID   ADX26534 standard; protein; 225 AA.
AC   ADX26534;
XX
DT   21-APR-2005   (first entry)
XX
DE   Discosoma DsRed RFP protein, seq id 12.
XX
KW   Phosphorylation; detection; red fluorescent protein; RFP.
OS   Discosoma sp.
XX
PN   US2005026234-A1.
XX
PD   03-FEB-2005.
XX
PF   28-MAY-2004; 2004US-00857622.
XX
PR   31-JAN-1996; 96US-00594575.
XX
PF   26-NOV-2003; 2003WO-RU000525.
XX
PR   26-DEC-2002; 2002US-04368572P.
PR   02-APR-2003; 2003US-0459679P.
XX
PA   (EVRO-) EVROGEN STOCK CO.
XX
PI   Barsova EV, Lukyanov SA;
XX
DR   WPI; 2004-525887/50.
XX
PT   New isolated nucleic acid molecule encoding a fluorescent protein, useful
PT   for labeling biomolecules, cell, or cell organelles, or for identifying
PT   expression of a gene in a biological specimen.
XX
PS   Example 1; Fig 1; 55pp; English.
XX
CC   The present invention relates to novel coding sequences (I) encoding
CC   green fluorescent proteins (II). The fluorescent protein coding sequences
CC   are isolated from an organism from phylum Arthropoda, from subclass
CC   Copepoda, or from family Pontellidae. The fluorescent protein coding
CC   sequences are useful for labelling biomolecules, cell, or cell
CC   organelles. They are also useful for identifying expression of a gene in
CC   a biological specimen or for generating transformants including
CC   transgenic organisms or site-specific gene modifications in cell lines.
CC   The fluorescent proteins are also useful in high throughput screening
CC   assays. The present sequence was used in a sequence alignment with the
CC   green fluorescent proteins of the invention.
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 8;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

Column 1

```
PR   31-JAN-1997;  97US-00792553.
PR   13-SEP-1999;  99US-00396003.
PR   24-MAY-2001;  2001US-00865291.
XX
PA   (VIOL/)  VIOLIN J D.
PA   (NEWT/)  NEWTON A C.
PA   (TSIE/)  TSIEN R Y.
PA   (ZHAN/)  ZHANG J.
XX
PI   Violin JD, Newton AC, Tsien RY, Zhang J;
XX
DR   WPI; 2005-141377/15.
DR   N-PSDB; ADX26533.
XX
PS   Disclosure; SEQ ID NO 12; 68pp; English.
XX
PT   New chimeric phosphorylation indicator
PT   protein, phosphoaminoacid binding domain, protein
PT   kinase C-phosphorylatable domain, useful for detecting protein kinase C
PT   or phosphates.
XX
CC   The invention relates to a chimeric phosphorylation indicator (CPI) (I)
CC   comprising in operative linkage, a first fluorescent protein,
CC   phosphoaminoacid binding domain with an FHA2 sequence defined in the
CC   specification, a protein kinase C (PKC)-phosphorylatable domain, and a
CC   second fluorescent protein. The first and second fluorescent proteins are
CC   chosen from green fluorescent protein (GFPs) and red fluorescent
CC   proteins (RFPs). The first and second fluorescent proteins exhibit a
CC   detectable resonance energy transfer, when the first fluorescent protein
CC   is excited. The PKC-phosphorylatable domain and phosphoaminoacid binding
CC   domain do not substantially emit light to excite the second fluorescent
CC   protein. (I) is useful for detecting a protein kinase C (PKC) or
CC   phosphates in a sample (such as biological sample e.g. cell, tissue
CC   sample, or their extract). (I) is useful for detecting a kinase
CC   inhibitor or phosphatase inhibitor. The method of the invention is
CC   adapted to high throughput analysis. The current sequence represents the
CC   amino acid sequence of Discosoma DsRed RFP.
XX
SQ   Sequence 225 AA;

Query Match             100.0%;  Score 1214;  DB 9;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 1e-127;
Matches  225;           Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSPPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSPPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 20
ABB08821
```
ID   ABB08821 standard; protein; 487 AA.
XX
AC   ABB08821;
XX
DT   28-MAY-2002  (first entry)
XX
DE   Autofluorescent fusion protein A SEQ ID NO 1.
XX
KW   Autofluorescent; fusion protein; proteolytic activity; fluorophore;
KW   protease; assay; protease cleavage.
```

Column 2

```
XX
OS   Synthetic.
XX
PN   WO200212543-A2.
XX
PD   14-FEB-2002.
XX
PF   07-AUG-2001; 2001WO-EP009112.
XX
PR   07-AUG-2000; 2000DE-01038382.
XX
PA   (DIRB-) DIRBVO BIOTECH AG.
XX
PI   Kuhlemann R, Koltermann A, Kettling U, Schwille P;
XX
PS   Example; Page 11; 35pp; German.
XX
CC   The invention relates to an autofluorescing fusion protein (I, ABB08821-
CC   ABB08823) comprising: (i) a first autofluorescing protein (Ia); (ii) a
CC   segment containing a protease cleavage site; and (iii) at least one
CC   different autofluorescing protein (Ib). Essentially no fluorescent energy
CC   transfer occurs between (Ia) and (Ib). (II) is used for detecting and
CC   quantifying protease (or protease-inhibitory) activity in liquid samples
CC   or cells, particularly for screening-based optimisation (or generation)
CC   of biomolecules with proteolytic activity. (I) can be prepared in
CC   cellular or cell-free systems and makes possible intracellular analysis
CC   of protease activity. Preparation of (I) does not require regioselective
CC   coupling of fluorophores to polypeptides and any selected protease
CC   cleavage site can be incorporated
XX
SQ   Sequence 487 AA;

Query Match             100.0%;  Score 1214;  DB 5;  Length 487;
Best Local Similarity   100.0%;  Pred. No. 3.2e-127;
Matches  225;           Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  263  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 322

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSPPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  323  LSPQFQYGSKVYVKHPADIPDYKKLSPPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 382

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  383  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 442

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  443  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 487
```

RESULT 21
ABB08822
```
ID   ABB08822 standard; protein; 506 AA.
XX
AC   ABB08822;
XX
DT   28-MAY-2002  (first entry)
XX
DE   Autofluorescent fusion protein B SEQ ID NO 2.
XX
KW   Autofluorescent; fusion protein; proteolytic activity; fluorophore;
KW   protease; assay; protease cleavage.
XX
OS   Synthetic.
```

Column 3

```
XX
XX
PT   New autofluorescent fusion protein, useful for determining proteins and
PT   protease-inhibiting activity, comprises two different proteins linked by
PT   protease cleavage site.
XX
DR   WPI; 2002-269094/31.
XX
```

RESULT 22

ABB08823

```
ID   ABB08823 standard; protein; 547 AA.
XX
AC   ABB08823;
XX
DT   28-MAY-2002  (first entry)
XX
DE   Autofluorescent fusion protein C SEQ ID NO 3.
XX
KW   Autofluorescent; fusion protein; proteolytic activity; fluorophore;
KW   protease; assay; protease cleavage.
XX
OS   Synthetic.
XX
PN   WO20021254 3-A2.
XX
PD   14-FEB-2002.
XX
PP   07-AUG-2001; 2001WO-EP009112.
XX
PR   07-AUG-2000; 2000DE-01038382.
XX
PA   (DIRE-) DIREVO BIOTECH AG.
XX
PI   Kuhlemann R, Koltermann A, Kettling U, Schwille P;
XX
DR   WPI; 2002-269094/31.
XX
PT   New autofluorescent fusion protein, useful for determining protease and
PT   protease-inhibiting activity, comprises two different proteins linked by
PT   protease cleavage site.
XX
PS   Claim 6; Page 11; 35pp; German.
XX
CC   The invention relates to an autofluorescing fusion protein (I, ABB08821-
CC   ABB08823) comprising: (ii) a first autofluorescing protein (Ia); (ii) a
CC   segment containing a protease cleavage site; and (iii) at least one
CC   different autofluorescing protein (Ib). Essentially no fluorescent energy
CC   transfer occurs between (Ia) and (Ib). (I) is used for detecting and
CC   quantifying protease (or protease-inhibitory) activity in liquid samples
CC   or cells, particularly for screening-based optimisation (or generation)
CC   of biomolecules with proteolytic activity. (I) can be prepared in
CC   cellular or cell-free systems and makes possible intracellular analysis
CC   of protease activity. Preparation of (I) does not require regioselective
CC   coupling of fluorophores to polypeptides and any selected protease
CC   cleavage site can be incorporated
XX
SQ   Sequence 506 AA;

Query Match             100.0%;  Score 1214;  DB 5;  Length 506;
Best Local Similarity   100.0%;
Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 341

Qy         61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 401

Qy        121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 461

Qy        181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
               |||||||||||||||||||||||||||||||||||||||||||||
Db        462  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 506
```

RESULT 23

AAE28920

```
ID   AAE28920 standard; protein; 225 AA.
XX
AC   AAE28920;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, E5 (S197T).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
KW   mutein.
XX
OS   Discosoma sp.
XX
PN   WO20021254 3-A2.
XX
PD   14-FEB-2002.
XX
PP   07-AUG-2001; 2001WO-EP009112.
XX
PR   07-AUG-2000; 2000DE-01038382.
XX
PA   (DIRE-) DIREVO BIOTECH AG.
XX
PI   Kuhlemann R, Koltermann A, Kettling U, Schwille P;
XX
DR   WPI; 2002-269094/31.
XX
PT   New autofluorescent fusion protein, useful for determining protease and
PT   protease-inhibiting activity, comprises two different proteins linked by
PT   protease cleavage site.
XX
PS   Claim 6; Page 11; 35pp; German.
XX
CC   The invention relates to an autofluorescing fusion protein (I, ABB08821-
CC   ABB08823) comprising: (ii) a first autofluorescing protein (Ia); (ii) a
CC   segment containing a protease cleavage site; and (iii) at least one
CC   different autofluorescing protein (Ib). Essentially no fluorescent energy
CC   transfer occurs between (Ia) and (Ib). (I) is used for detecting and
CC   quantifying protease (or protease-inhibitory) activity in liquid samples
CC   or cells, particularly for screening-based optimisation (or generation)
CC   of biomolecules with proteolytic activity. (I) can be prepared in
CC   cellular or cell-free systems and makes possible intracellular analysis
CC   of protease activity. Preparation of (I) does not require regioselective
CC   coupling of fluorophores to polypeptides and any selected protease
CC   cleavage site can be incorporated
XX
SQ   Sequence 547 AA;

Query Match             100.0%;  Score 1214;  DB 5;  Length 547;
Best Local Similarity   100.0%;
Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 382

Qy         61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 442

Qy        121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 502

Qy        181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
               |||||||||||||||||||||||||||||||||||||||||||||
Db        503  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 547
```

OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 197
FT   /note= "Wild-type Ser substituted with Thr"
XX
PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-0006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescence timers. The present
CC   sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAB28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   99.8%; Score 1211; DB 5; Length 225;
Best Local Similarity   99.6%;
Matches 224; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||:||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 24
ABW00937
ID   ABW00937 standard; protein; 225 AA.
XX
AC   ABW00937;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), S197T.

XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 197
FT   /note= "Wild-type Ser is substituted with Thr"
XX
PN   US20031709111-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
DR   WPI; 2003-802418/75.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PS   Example 2; Page; 0pp; English.
XX
CC   Fluorescent proteins containing a mutation that reduces or eliminates its
CC   ability to oligomerize which gives more reliable fluorescence resonance
CC   energy transfer results and are useful to detect molecule interaction,
CC   enzymes, or sample pH.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match   99.8%; Score 1211; DB 7; Length 225;
Best Local Similarity   99.6%;
Matches 224; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||:||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 25
ABW00929
ID   ABW00929 standard; protein; 225 AA.
XX
AC   ABW00929;

RESULT 26

Left column:

```
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K83R.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 83
FT                   /note= "wild-type Lys is substituted with Arg"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 2; Page; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match             99.8%;  Score 1211;  DB 7;  Length 225;
Best Local Similarity   99.6%;  Pred. No. 2.2e-127;
Matches  224;  Conservative  1;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

Right column:

```
ABW00938
ID   ABW00938 standard; protein; 225 AA.
XX
AC   ABW00938;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K70R.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 70
FT                   /note= "wild-type Lys is substituted with Arg"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 2; Page; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match             99.8%;  Score 1211;  DB 7;  Length 225;
Best Local Similarity   99.6%;  Pred. No. 2.2e-127;
Matches  224;  Conservative  1;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSRYTVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

Db
181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 27
ID   ADH34498 standard; protein; 225 AA.
XX
AC   ADH34498;
XX
DT   11-MAR-2004 (first entry)
XX
DE   Discosoma sp. DsRed mutant S148A.
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   red fluorescence activated cell sorting; fluorescent timer;
KW   red fluorescent protein; DsRed; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 146
FT   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT   residue 148"
XX
PN   WO20030057833-A2.
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.
XX
PR   26-DEC-2001; 2001US-0343128P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein of the invention; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants are useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can be useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and

CC   are also useful as sunscreens and selective filters. Chromoprotein
CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g., transgenic
CC   animals), in fluorescence assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g., green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence
CC   represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC   in an example of the invention. The present sequence is not shown in the
CC   specification, but was derived from the wild-type DsRed sequence
CC   (ADH34489) shown in Fig 1 and the information provided on page 42.
XX
SQ   Sequence 225 AA;

Query Match   99.8%; Score 1211; DB 7; Length 225;
Best Local Similarity   99.6%; Pred. No. 2.2e-127;
Matches 224; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEAATERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 28
ID   ADH34499 standard; protein; 225 AA.
XX
AC   ADH34499;
XX
DT   11-MAR-2004 (first entry)
XX
DE   Discosoma sp. DsRed mutant S203A.
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   red fluorescence activated cell sorting; fluorescent timer;
KW   red fluorescent protein; DsRed; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 197
FT   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT   residue 203"
XX
PN   WO20030057833-A2.
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.

PR   26-DEC-2001; 2001US-0343128P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The
CC   invention is derived from a Cnidarian species, preferably a non-bioluminescent
CC   Cnidarian species, and most preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants are useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can be useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   having chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as sunscreens and selective filters. Chromoprotein
CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g. transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g. green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence
CC   represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC   in an example of the invention. The present sequence is not shown in the
CC   specification, but was derived from the wild-type DsRed sequence
CC   (ADH3489) shown in Fig 1 and the information provided on page 42.
XX
SQ   Sequence 225 AA;

Qy   Sequence 225 AA;

Query Match   99.8%;   Score 1211; DB 7; Length 225;
Best Local Similarity   99.6%;   Pred. No. 2.2e-127;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||| ||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 29
AAE28919
ID   AAE28919 standard; protein; 225 AA.
XX
AC   AAE28919;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, E5 (V105A).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
KW   mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference   105
FT   /note= "wild-type Val substituted with Ala"
XX
PN   WO200268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo-/fluorescent proteins and their mutants. Chromo-/fluoro proteins
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is a Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   99.7%;   Score 1210; DB 5; Length 225;
Best Local Similarity   99.6%;   Pred. No. 2.9e-127;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 30
ABW00930

ID   ABW00930 standard; protein; 225 AA.
AC   ABW00930;
DT   15-JAN-2004 (first entry)
DE   Discosoma sp. red fluorescent protein (RFP), K83E.
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
OS   Discosoma sp.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   Misc-difference 83
FT   /note= "wild-type Lys is substituted with Glu"

PI   Tsien RY, Zacharias DA, Baird GS;
PR   WPI; 2003-802418/75.

PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.

PN   US2003170911-A1.
PD   11-SEP-2003.

PF   26-FEB-2001; 2001US-0079430B.
PR   26-FEB-2001; 2001US-0079430B.

PS   Example 2; Page; Opp; English.

PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.

CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)

SQ   Sequence 225 AA;

Query Match   99.7%; Score 1210; DB 7; Length 225;
 .   Best Local Similarity   99.6%; Pred. No. 2.9e-127;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
Db   1 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVWNFEDGGVVTVTQDSSLQDGCFIY 120
Qy 61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVWNFEDGGVVTVTQDSSLQDGCFIY 120
Db 61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVWNFEDGGVATVTQDSSLQDGCFIY 120
Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 31
AAG65510

ID   AAG65510 standard; protein; 226 AA.
AC   AAG65510;
DT   30-NOV-2001 (first entry)
DE   Anthozoan red fluorescent protein synthetic mutant sequence.
KW   Fluorescent protein; Anthozoan; fluorescence; marker; FRET; mutant.
OS   Synthetic.

PI   Nelson D, Zamaira E, Tsien R;
PA   (AURO-) AURORA BIOSCIENCES CORP.

PN   WO20016291-A1.
PD   30-AUG-2001.

PF   13-FEB-2001; 2001WO-US004625.
PR   23-FEB-2000; 2000US-0184732P.

PS   Disclosure; Page 87; 90pp; English.

PT   Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
PT   functional red fluorescent proteins, and the encoding nucleic acid, with
PT   key mutations for improving the proteins function.

DR   WPI; 2001-557704/62.
DR   N-PSDB; AAH47656.

CC   The invention provides a nucleic acid encoding functional red fluorescent
CC   protein (II) that differs from the sequence of an Anthozoan red
CC   fluorescent protein by at least one amino acid substitution, and with
CC   different fluorescent properties. The red fluorescent protein of the
CC   invention can be expressed by standard recombinant methodology. (II) are
CC   used a fluorescent markers and FRET partners. It is used for identifying
CC   protein-protein interactions. (II) is also suitable for multiplexed
CC   fluorescent analysis and FRET-based applications using existing Aequorea
CC   fluorescent proteins. (II) has improved brightness, reduced spectral
CC   cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC   The key mutations in the encoding nucleic acids provide improved folding,
CC   brightness, and create (II) with sharper, more defined excitation and
CC   emission peaks when expressed in mammalian cells. The present sequence
CC   represents the amino acid sequence of an improved synthetic mutant of an
CC   anthozoan fluorescent protein

SQ   Sequence 226 AA;

Query Match   99.7%; Score 1210; DB 4; Length 226;
 Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
Qy 61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVWNFEDGGVVTVTQDSSLQDGCFIY 120
Db 61 LSPQFQYGSKVYVKHPADIPDYEKLSFPEGFKWERVWNFEDGGVATVTQDSSLQDGCFIY 120
Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Qy Beet Local Similarity 99.6%; Pred. No. 2.9e-127; Indels 0; Gaps 0;
    Matches 224; Conservative 1; Mismatches 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKYYVKHPADIDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKYYVKHPADIDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db 182 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 32
ID   ADE24109
AC   ADE24109;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Discosoma sp DsRed1 reporter protein.
XX
KW   site-specific DNA recombination; Cre recombinase;
KW   protein-transduction domain; gene activation; gene inactivation;
KW   chromosomal translocation; reporter protein.
XX
OS   Discosoma sp.
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
PN   WO20030709l1-A2.
XX
PD   28-AUG-2003.
XX
PF   19-FEB-2003; 2003WO-EP001680.
PR   21-FEB-2002; 2002DE-01007313.
PR   16-JUL-2002; 2002DE-01032196.
XX
PA   (VISI-) VISION 7 GMBH.
XX
PI   Baum C, Will E, Ostertag W, Klump H, Schiedlmeier B;
XX
XX
DR   WPI; 2003-767353/72.
XX
PT   In vitro or in vivo site-specific DNA recombination, useful e.g. for gene
PT   inactivation, using Cre recombinase that lacks heterologous protein-
PT   transduction domain.
XX
PS   Claim 22; SEQ ID NO 6; 84pp; German.
XX
XX
CC   This invention describes a novel method of site-specific DNA
CC   recombination in eukaryotic cells in vitro comprising using a
CC   bacteriophage P1 Cre recombinase that lacks heterologous protein-
CC   transduction domain. The invention also contains a reporter system for
CC   detecting site-specific DNA recombination in eukaryotic cells.
CC   Recombination with Cre recombinase lacking heterologous protein-
CC   transduction domains is used for therapeutic site-specific recombination
CC   in eukaryotic cells, in vivo or in vitro, e.g. for partial inactivation
CC   or activation of genes, also to create a chromosomal translocation. The
CC   modified recombinase can enter intact (not electrically or chemically
CC   treated) cells when added to culture medium and catalyzes recombination
CC   events in the nucleus with recombination rate over 50%, after a single
CC   application. The Cre-recombinase does not need to be supplied from
CC   nucleic acid introduced into the cell, so its effect is immediate but
CC   short-lived, limiting cytotoxic effects; eliminating the risk of non-
CC   specific integration of DNA into the genome and providing quick results.

CC   This sequence represents the Discosoma sp reporter protein DsRed1 used in
CC   the method described in the invention.
XX
SQ   Sequence 242 AA;

    Query Match   99.7%; Score 1210; DB 7; Length 242;
    Best Local Similarity 99.6%; Pred. No. 3.2e-127; DB 7; Length 242;
    Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 74

Qy   61 LSPQFQYGSKYYVKHPADIDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 LSPQFQYGSKYYVKHPADIDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY 134

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 194

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db 195 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 239

RESULT 33
ID   ADL18132
AC   ADL18132;
XX
DT   06-MAY-2004 (first entry)
XX
DE   RFP:PS(NIa protease):AtOEP7:GFP fusion protein SEQ ID NO:52.
XX
KW   chimeric protein; signal protein; trafficking signal targeting;
KW   proteolytic cleavage site; protease; protease inhibitor.
XX
OS   Arabidopsis thaliana.
OS   Tobacco vein mottling virus.
OS   Synthetic.
XX
XX
PN   WO2003014381-A1.
XX
PD   20-FEB-2003.
XX
PF   08-AUG-2002; 2002WO-KR001515.
PR   10-AUG-2001; 2001KR-00048123.
XX
PA   (AHRA-) AHRAM BIOSYSTEMS INC.
XX
PI   Hwang I, Kim DH, Lee YJ;
XX
XX
DR   WPI; 2003-265596/25.
DR   N-PSDB; ADL18131.
XX
PT   New chimeric protein, useful for detecting protease inhibitors inside the
PT   cell or tissue.
XX
PS   Example 2; SEQ ID NO 52; 214pp; English.
XX
XX
CC   The present invention describes a chimeric protein comprising at least
CC   one signal protein that has a trafficking signal targeting to a
CC   subcellular organelle and at least one proteolytic cleavage site for a
CC   protease. The chimeric protein is constructed, so that: (a) the
CC   trafficking signals and all the signal proteins are inactivated by linking
CC   the proteolytic site or a signal masking protein through the proteolytic
CC   site to the N-or C- terminus of the signal proteins, and so the chimeric
CC   protein is present in cytosol; (b) the trafficking signal of at least one
CC   signal protein is activated when the proteolytic cleavage site is cleaved
CC   by the protease, and as a result at least one fragment protein that

CC includes the activated signal protein is a transported to a subcellular
CC organelle; and (c) the chimeric protein is labelled with at least one
CC fluorescent protein and the position and intensity distribution of the
CC fluorescent label signal in the cell is altered depending on the cleavage
CC by the protease. Also described: (1) a recombinant gene comprising a
CC nucleic acid sequence encoding the chimeric protein which is constructed
CC to express the chimeric protein in a cell; (2) a cell transformed with
CC the recombinant gene or vector; (3) analysing the activity of a protease
CC in vivo; (4) screening protease inhibitors in vivo; (5) a system for
CC detecting a protease inside a cell; (6) a nucleic acid comprising the
CC sequence encoding the chimeric protein for detecting protease activity in
CC a cell; (7) a vector comprising the nucleic acid; (8) a kit for detecting
CC a protease inside a cell comprising the chimeric protein or the vector;
CC (9) detecting a protease inside a cell comprising the chimeric protein;
CC protease inhibitor in vivo. The chimeric protein is useful for detecting
CC protease inhibitors inside the cell or tissue. The present sequence
CC represents a fusion protein, which is used in the exemplification of the
CC present invention.
CC
SQ Sequence 545 AA;

Query Match   Best Local Similarity   99.7%;
                Conservative 99.6%;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

                Score 1210; DB 7; Length 545;
                Pred. No. 1.1e-126;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   61

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 182 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 34
ID   ADL18156 standard; protein; 548 AA.
XX
AC   ADL18156;
XX
DT   06-MAY-2004 (first entry)
XX
DE   RFP:PS(HIV-1 protease):AtOEP7:GFP fusion protein SEQ ID NO:76.
XX
KW   chimeric protein; signal protein; trafficking signal targeting;
KW   proteolytic cleavage site; protease; protease inhibitor.
XX
OS   Arabidopsis thaliana.
OS   Human immunodeficiency virus 1.
OS   Synthetic.
XX
PN   WO2003014381-A1.
XX
PD   20-FEB-2003.
XX
PF   08-AUG-2002; 2002WO-KR001515.
PR   10-AUG-2001; 2001KR-00048123.
XX
PA   (AHRA-) AHRAM BIOSYSTEMS INC.
XX
PI   Hwang I, Kim DH, Lee YJ;
XX
DR   WPI; 2003-256596/25.
DR   N-PSDB; ADL18155.

XX
PT   New chimeric protein, useful for detecting protease inhibitors inside the
PT   cell or tissue.
XX
PS   Example 3; SEQ ID NO 76; 214pp; English.
XX
CC   The present invention describes a chimeric protein comprising at least
CC   one signal protein that has a trafficking signal targeting to a
CC   subcellular organelle and at least one proteolytic cleavage site for a
CC   protease. The chimeric protein is constructed, so that: (a) the
CC   trafficking signals of all the signal proteins are inactivated by linking
CC   the protease or a signal masking protein, through the proteolytic
CC   site to the N-or C- terminus of the signal proteins, and so the chimeric
CC   signal protein is activated when the proteolytic cleavage site is cleaved
CC   by the protease, and as a result at least one fragment protein that
CC   includes the activated signal protein is transported to a subcellular
CC   organelle; and (c) the chimeric protein is labelled with at least one
CC   fluorescent protein and the position and intensity distribution of the
CC   fluorescent label signal in the cell is altered depending on the cleavage
CC   by the protease. Also described: (1) a recombinant gene comprising a
CC   nucleic acid sequence encoding the chimeric protein in a cell; (2) a cell
CC   to express the chimeric protein in a cell; (3) analysing the activity of a
CC   the recombinant gene or vector; (3) analysing the activity of a protease
CC   in vivo; (4) screening protease inhibitors in vivo; (5) a system for
CC   detecting a protease inside a cell; (6) a nucleic acid comprising the
CC   sequence encoding the chimeric protein for detecting protease activity in
CC   a cell; (7) a vector comprising the nucleic acid; (8) a kit for detecting
CC   a protease inside a cell comprising the chimeric protein or the vector;
CC   (9) detecting a protease inside a cell or tissue; and (10) detecting a
CC   protease inhibitor in vivo. The chimeric protein is useful for detecting
CC   protease inhibitors inside the cell or tissue. The present sequence
CC   represents a fusion protein, which is used in the exemplification of the
CC   present invention.
CC
SQ Sequence 548 AA;

Query Match   Best Local Similarity   99.7%;
                Conservative 99.6%;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

                Score 1210; DB 7; Length 548;
                Pred. No. 1.1e-126;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   61

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 182 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 35
ID   AAE28922 standard; protein; 225 AA.
XX
AC   AAE28922;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, E8 (N42H).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;

KW   mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 42
FT   /note= "wild-type Asn substituted with His"
XX
XX
PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S; Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
XX
PT   New nucleic acid encoding a non-aggregating Cnidarian chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   applications.
XX
PS   Disclosure; Page; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   99.6%;   Score 1209; DB 5; Length 225;
Best Local Similarity 99.6%; Pred. No. 3.8e-127;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHHTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 36
ID   ABW00936
AC   ABW00936;
XX
XX

DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), Y120H.
XX
KW   Fluorescent protein; resonance energy transfer; Y120H.
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 120
FT   /note= "wild-type Tyr is substituted with His"
XX
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   Tsien RY, Zacharias DA, Baird GS;
XX
PI   Tsien R Y.;
PI   (TSIE/) TSIEN R Y.;
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
DR   WPI; 2003-802418/75.
XX
PS   Example 2; Page; 0pp; English.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match   99.6%;   Score 1209; DB 7; Length 225;
Best Local Similarity 99.6%; Pred. No. 3.8e-127;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHHTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIH 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 37
ABW00931

ID   ABW00931 standard; protein; 225 AA.
XX
AC   ABW00931;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K83N.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 83
FT   /note= "Wild-type Lys is substituted with Asn"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 2; Page; Opp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match   99.6%; Score 1209; DB 7; Length 225;
Best Local Similarity   99.6%; Pred. No. 3.8e-127;
Matches 224; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
               |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 38
ID   ABB08835 standard; protein; 226 AA.
XX
AC   ABB08835;
XX
DT   29-MAY-2002 (first entry)
XX
DE   Yeast optimised RFP SEQ ID NO 20.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology.
XX
OS   Anthozoa.
XX
PN   DE20001395-U1.
XX
PD   15-MAR-2001.
XX
PF   27-JAN-2000; 2000DE-02001395.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
PI   WPI; 2002-228394/29.
DR   N-PSDB; ABA95906.
XX
PT   New DNA encoding red fluorescent protein, useful as marker in
PT   biotechnology, has sequence optimized for expression in eukaryotes,
PT   especially yeast or plants.
XX
PS   Disclosure; Page 15; 19pp; German.
XX
CC   The invention relates to DNA (I) containing either sequence ABA95906 or
CC   sequence ABA95905 encoding a yeast optimised Red Fluorescent Protein
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially dicotyledonous plants including Nicotiana tabacum or
CC   plants, especially yeast, especially Saccharomyces cerevisiae and
CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern
CC   biotechnology. (I) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio. The present
CC   sequence is that of the yeast optimised RFP
XX
SQ   Sequence 226 AA;

Query Match   99.6%; Score 1209; DB 5; Length 226;
Best Local Similarity 100.0%; Pred. No. 3.8e-127;
Matches 224; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2 RSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDIL 61
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3 RSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDIL 62

Qy   62 SPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYK 121
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 SPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYK 122

Qy   122 VKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIY 181
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 VKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIY 182

Qy   182 MAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
               ||||||||||||||||||||||||||||||||||||||||||||
Db   183 MAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 39
ID   ABW00932 standard; protein; 225 AA.
XX
AC   ABW00932;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K83P.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 83
FT   /note= "wild-type Lys is substituted with Pro"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 2; Page; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match   99.5%; Score 1208; DB 7; Length 225;
Best Local Similarity   99.6%; Pred. No. 4.9e-127;
Matches 224; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 40
ID   ABW00935 standard; protein; 225 AA.
XX
AC   ABW00935;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K83M.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 83
FT   /note= "wild-type Lys is substituted with Met"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
DR   WPI; 2003-802418/75.
XX
PA   (TSIE/) TSIEN R Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 2; Page; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Query Match   99.5%; Score 1208; DB 7; Length 225;
Best Local Similarity   99.6%; Pred. No. 4.9e-127;
Matches 224; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYMCLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

```
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 41
ID   ABW00939
AC   ABW00939;
XX
ID   ABW00939 standard; protein; 225 AA.
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K70M.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FT   Misc-difference  70
FT                  /note= "wild-type Lys is substituted with Met"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PR   26-FEB-2001; 2001US-0079430B.
XX
PR   26-FEB-2001; 2001US-0079430B.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
PA   (TSIE/) TSIEN R.Y.
PA   (ZACH/) ZACHARIAS D.A.
PA   (BAIR/) BAIRD G.S.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
DR   WPI; 2003-802418/75.
XX
PS   Example 2; Page; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Qy   Query Match  99.5%;  Score 1208;  DB 7;  Length 225;
     Best Local Similarity  99.6%;  Pred. No. 4.9e-127;
     Matches  224; Conservative  0; Mismatches  1; Indels  0; Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
```

```
Db   61 LSPQFQYGSMVYKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 180

RESULT 42
ID   ADH34500
AC   ADH34500;
XX
ID   ADH34500 standard; protein; 225 AA.
XX
DT   11-MAR-2004 (first entry)
XX
DE   Discosoma sp. DsRed mutant S148A/S203A.
XX
FH   Key            Location/Qualifiers
FT   Misc-difference 146
FT                  /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT                  residue 148"
FT   Misc-difference 197
FT                  /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT                  residue 203"
XX
OS   Synthetic.
OS   Discosoma sp.
XX
PN   WO200305783-A2.
XX
PD   17-JUL-2003.
XX
PR   23-DEC-2002; 2002WO-US041418.
XX
PR   26-DEC-2001; 2001US-03431288P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins. The
CC   fluorescent proteins (FP) and nucleic acids encoding them. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoprotein being able to be converted into fluorescent proteins in a
```

CC similar manner. The invention also relates to expression constructs,
CC vectors, host cells and host cell progeny comprising a nucleic acid of
CC the invention; the recombinant production of an interconverted
CC chromoprotein or fluorescent protein mutant; and antibodies specific for
CC interconverted mutant proteins of the invention. The interconverted
CC chromoprotein or fluorescent protein mutant proteins having chromoprotein or
CC fluorescent protein. Fluorescent protein mutants having chromoprotein
CC activity can useful in any application that employs a chromoprotein or
CC fluorescent protein, for example, food
CC compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC with chromoprotein activity are also useful as labels in biological
CC analyte detection assays, as selectable markers in recombinant DNA
CC applications (e.g. the production of transgenic cells and organisms), and
CC are also useful as sunscreens and selective filters. Chromoprotein
CC mutants having fluorescent protein activity useful in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC changes in multicellular reorganisation and migration, as second
CC messenger detectors, as in vivo markers in animals (e.g.; transgenic
CC animals), in fluorescence activated cell sorting applications, in
CC protease cleavage assays, and in assays to determine the phospholipid
CC composition in biological membranes. Proteins with fluorescent protein
CC activity can also be used as fluorescent timers, where the switch of one
CC fluorescent colour to another (e.g.; green to red) is concomitant with
CC the ageing of the protein and is useful for determination of the
CC activation or deactivation of gene expression. The present sequence
CC represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC in an example of the invention. The present sequence is not shown in the
CC specification, but was derived from the wild-type DsRed sequence
CC (ADH34489) shown in Fig 1 and the information provided on page 42.

XX

SQ Sequence 225 AA;

Query Match   99.5%; Score 1208; DB 7; Length 225;
Best Local Similarity   99.1%;
Matches 223; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           ||||||||||||||:|||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

XX

RESULT 43
AC AAE28921
ID AAE28921 standard; protein; 225 AA.
XX
AC AAE28921;
XX
DT 27-DEC-2002 (first entry)
XX
DE Discosoma sp. drFP583 (NFP-6) mutant protein, E5 (V105A+S197T).
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
KW mutein.
XX
OS Discosoma sp.
OS Synthetic.
XX

FH Key   Location/Qualifiers
FT Misc-difference 105
FT   /note= "Wild-type Val substituted with Ala"
FT Misc-difference 217
FT   /note= "Wild-type Ser substituted with Thr"
XX
PN WO20026459-A2.
XX
PD 06-SEP-2002.
XX
PF 20-FEB-2002; 2002WO-US005749.
XX
PR 21-FEB-2001; 2001US-0270983P.
PR 04-DEC-2001; 2001US-00006922.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
DR WPI; 2002-691654/74.
XX
CC New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
CC protein. Chromo/fluoro proteins are
CC useful in analyte detection assays, in screening assays, as
CC recombinant DNA applications, as biosensors in
CC prokaryotic and eukaryotic cells, in fluorescence activated cell sorting applications,
CC resonance energy transfer (FRET) applications, as sunscreens or filters, in fluorescence
CC chromo/fluorescence proteins and their mutants. Chromo/fluoro proteins are
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC invention. Note: This sequence is not shown in the specification, but is
CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC ID NO:8 (AAE28833) in page 70-71 of the specification
CC analyte detection assays or fluorescence activated cell sorting
CC applications.
XX

The invention relates to nucleic acid molecules encoding non-aggregating

Disclosure; Page; 80pp; English.

XX

SQ Sequence 225 AA;

Query Match   99.4%; Score 1207; DB 5; Length 225;
Best Local Similarity   99.1%;
Matches 223; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERTEGRHHLFL 225

XX

RESULT 44
AC AAE17541
ID AAE17541 standard; protein; 225 AA.
XX
AC AAE17541;
XX
DT 22-APR-2002 (first entry)
XX
DE Discosoma sp. fluorescent protein E5.

Qy .

Db 1   MRSSKNVIKEFMRFKVRMEGTVNGHFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1   MRSSKNVIKEFMRFKVRMEGTVNGHFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy 61 LSPQFQYGSKYVKVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Matches 223; Conservative   1; Mismatches   1; Indels   0; Gaps   0;
Best Local Similarity 99.4%;   Pred. No. 6.3e-127;
Query Match 99.4%;   Score 1207; DB 5;   Length 225;

SQ Sequence 225 AA;

XX
DR drPS83
CC fluorescent protein derived from humanised wild-type Anthozoa protein
CC modifications in cell lines. The present sequence is Discosoma sp. E5
CC to generate transgenic, non-human plants or animals or site-specific gene
CC assays, and as second messenger detectors. The nucleic acids can be used
CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC techniques are employed, as detectable labels, as selectable markers, as
CC timer proteins may further be used to investigations where photobleaching
CC membrane junctions and extracellular matrix components. The fluorescent
CC in distinguishing between newly formed and pre-existing structures, e.g.
CC visualisation of newly synthesised proteins and accumulated proteins, and
CC fate during development and organ remodelling, in spatial and temporal
CC aspects of the activity of a regulatory element, for determining cell
CC protein trafficking, or protein stability, to investigate temporal
CC activity, monitoring intracellular protein movement or translocation,
CC promoter activity, as a reporter to serve as a read-out of promoter
CC cellular differentiation, in response to a drug or other inducer of
CC expression during development of a multicellular organism or during
CC The fluorescent timer proteins are also useful for assessing gene
CC enriching a population of cells comprising a fluorescent timer protein.
CC identifying an agent that modulates the activity of a promoter and in
CC monitoring the activity of a promoter, determining the age of a protein,
CC to a second wavelength. The fluorescent timer proteins are useful in
CC spectrum that changes over time after synthesis from a first wavelength
CC The invention relates to a fluorescent timer protein having an emission
XX
PT Claim 5; Fig 2; 89pp; English.
XX
PT or stability.
PT for monitoring intracellular protein movement, translocation, trafficking
PT changes over time from a first wavelength to a second wavelength, useful
PT New fluorescent timer proteins comprising an emission spectrum that
FH
XX
DR N-PSDB; AAD28208.
DR WPI; 2002-154595/20.
XX
PI Fradkov AF, Terskikh A;
XX
PA (CLON-) CLONTECH LAB INC.
XX
PR 14-JUN-2000; 2000US-0211607P.
PF 13-JUN-2001; 2001WO-US019097.
PD 20-DEC-2001.
XX
PN WO200196373-A2.
XX
FT /note= "Wild type Ser substituted with Thr"
FT Misc-difference 197
FT /note= "Wild type Val substituted with Ala"
    Misc-difference 105
    Location/Qualifiers
FH Key
XX
OS Discosoma sp.
XX
KW protein age; E5.
KW promoter activity; gene expression; transgenic plant; gene modification;
KW Fluorescent timer protein; protein movement; translocation; trafficking;
XX

Qy 61 LSPQFQYGSKYVKVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||

Db 181 YMAKKPVQLPGYYYVDYKLDITSHNEDYTIVEQYERTEGRHHLFL 225

SQ Sequence 225 AA;

XX
OS Synthetic.
OS Discosoma sp.
XX
KW mutein.
KW fluorescence; mutagenesis; red fluorescent protein; protein interaction;
XX
DE Discosoma sp. red fluorescent protein (RFP) mutant I125R.
XX
DT 05-MAY-2005 (first entry)
XX
AC ADY51734;
XX
ID ADY51734 standard; protein; 225 AA.
ADY51734

RESULT 45

Db 181 YMAKKPVQLPGYYYVDYKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 61 LSPQFQYGSKYVKVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120

Db

XX
FT /note= "Wild type Ile substituted by Arg"
FT Misc-difference 125
    Location/Qualifiers
FH Key
XX
OS Synthetic.
OS Discosoma sp.
XX

PN WO200268605-A2.
XX
PD 06-SEP-2002.
XX
PS Claim 11; Page; 117pp; English.
XX
PT biology.
PT oligomerize, useful for making better and new assays for molecular
PT mutation that reduces or eliminates the ability of the protein to
PT New non-oligomerizing fluorescent protein containing at least one
XX
DR WPI; 2002-713372/77.
XX
PI Tsien RY, Baird GS, Campbell RE, Zacharias DA;
XX
PA (REGC) UNIV CALIFORNIA.
XX
PR 24-MAY-2001; 2001US-00865538.
PR 26-FEB-2001; 2001US-00794308.
XX
PF 26-FEB-2002; 2002WO-US006063.
XX

CC information provided in the claims.
CC not given in the invention but it is derived from SeqID 12 and
CC fluorescent protein (RFP) mutant of the invention. NOTE: This sequence is
CC of molecular biology. This polypeptide sequence is the Discosoma sp. red
CC they may be used to identify a specific interaction of molecules, in the field
CC regulates the activity of an expression control sequence. Furthermore,
CC sample, or whether the sample contains an enzyme, molecule or agent that
CC proteins can be used in methods and compositions to determine the pH of a
CC protein linked to at least one protein of interest. As such, these fusion
CC provides a fusion protein that comprises a non-oligomerizing fluorescent
CC derived from naturally occurring green or red fluorescent proteins and
CC to oligomerize. The present invention describes fluorescent proteins
CC fluorescent protein that reduces or eliminates the ability of the protein
CC Specifically, it refers to the presence of at least one mutation in the
CC This invention relates to a novel non-oligomerizing fluorescent protein.
XX

SQ Sequence 225 AA;

Query Match 99.4%;   Score 1207; DB 5;   Length 225;

RESULT 46

Qy Best Local Similarity 99.6%; Pred. No. 6.3e-127; 1;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 46
ABW00941
ID   ABW00941 standard; protein; 225 AA.
XX
AC   ABW00941;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), I125R.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 125
FT   /note= "Wild-type Ile is substituted with Arg"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R.Y.
PA   (ZACH/) ZACHARIAS D.A.
PA   (BAIR/) BAIRD G.S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 3; Page; Opp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)

RESULT 47

XX
SQ   Sequence 225 AA;
XX

Query Match 99.4%; Score 1207; DB 7; Length 225;
Best Local Similarity 99.4%; Pred. No. 6.3e-127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Matches 224;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 47
ABW00940
ID   ABW00940 standard; protein; 225 AA.
XX
AC   ABW00940;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), I125K.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 125
FT   /note= "Wild-type Ile is substituted with Lys"
XX
PN   US2003170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PA   (TSIE/) TSIEN R.Y.
PA   (ZACH/) ZACHARIAS D.A.
PA   (BAIR/) BAIRD G.S.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
XX
DR   WPI; 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
XX
PS   Example 3; Page; Opp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The

CC present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC Note: This sequence is not shown in the specification, however this
CC sequence is constructed based on Discosoma sp. wild-type RFP protein
CC shown in page 30-31 (ABW00918)
XX
SQ Sequence 225 AA;

Query Match   99.4%; Score 1207; DB 7; Length 225;
Best Local Similarity 99.6%; Pred. No. 6.3e-127;
Matches 224; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 48

ID ADL46222 standard; protein; 225 AA.
XX
AC ADL46222;
XX
DT 20-MAY-2004 (first entry)
XX
DE Discosoma red fluorescent (DsRed) protein with I125R mutation.
XX
KW red fluorescent protein; DsRed; fluorescence; red wavelength;
KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS Discosoma sp.
XX
PN WO200308646-A1.
XX
PD 23-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010879.
XX
PR 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-0020208.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Campbell RE, Baird GS;
XX
DR WPI; 2003-845265/78.
XX
PS Disclosure; SEQ ID NO 20; 166pp; English.
XX
CC The invention relates to a polynucleotide sequence encoding a Discosoma
CC red fluorescent protein (DsRed) variant having a reduced propensity to
CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC sequence, where the substitutions result in reduced propensity of the
CC DsRed variant to form tetramers and where the variant displays detectable
CC fluorescence of at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the DsRed
CC protein with an I125R mutation.
XX
SQ Sequence 225 AA;

Query Match   99.4%; Score 1207; DB 7; Length 225;
Best Local Similarity 99.6%; Pred. No. 6.3e-127;
Matches 224; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 49

ID ABW00933 standard; protein; 225 AA.
XX
AC ABW00933;
XX
DT 15-JAN-2004 (first entry)
XX
DE Discosoma sp. red fluorescent protein (RFP), K83F.
XX
KW fluorescent protein; resonance energy transfer; pH; detection;
KW red fluorescent protein; RFP; mutant; mutein.
XX
OS Discosoma sp.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Misc-difference 83
FT   /note= "Wild-type Lys is substituted with Phe"
XX
PN US2003170911-A1.
XX
PD 11-SEP-2003.
XX
PF 26-FEB-2001; 2001US-00794308.
XX
PR 26-FEB-2001; 2001US-00794308.
XX
PA (TSIE/) TSIEN R Y.
PA (ZACH/) ZACHARIAS D A.
PA (BAIR/) BAIRD G S.
XX
PI Tsien RY, Zacharias DA, Baird GS;
XX
DR WPI; 2003-802418/75.
XX
PS Example 2; Page; 0pp; English.
XX
CC The invention relates to a non-oligomerising fluorescent protein
CC Fluorescent proteins containing a mutation that reduces or eliminates its
CC ability to oligomerize which gives more reliable fluorescence resonance
CC energy transfer results and are useful to detect molecule interaction,
CC enzymes, or sample pH.

```
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Qy   Query Match   99.3%;  Score 1206;  DB 7;  Length 225;
Best Local Similarity   99.6%;  Pred. No. 8.2e-127;
Matches 224;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 50
ID   ABW00934 standard; protein; 225 AA.
XX
AC   ABW00934;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP), K83W.
XX
KW   Fluorescent protein; resonance energy transfer; pH; detection;
KW   red fluorescent protein; RFP; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  83
FT                    /note= "wild-type Lys is substituted with Trp"
XX
PN   US20030170911-A1.
XX
PD   11-SEP-2003.
XX
PF   26-FEB-2001; 2001US-00794308.
XX
PR   26-FEB-2001; 2001US-00794308.
XX
PI   Tsien RY, Zacharias DA, Baird GS;
PA   (TSIE/) TSIEN R. Y.
PA   (ZACH/) ZACHARIAS D A.
PA   (BAIR/) BAIRD G S.
XX
DR   WPI, 2003-802418/75.
XX
PT   Fluorescent proteins containing a mutation that reduces or eliminates its
PT   ability to oligomerize which gives more reliable fluorescence resonance
PT   energy transfer results and are useful to detect molecule interaction,
PT   enzymes, or sample pH.
```

```
XX
PS   Example 2; Page; 0pp; English.
XX
CC   The invention relates to a non-oligomerising fluorescent protein
CC   containing a mutation that reduces or eliminates its ability to
CC   oligomerise. The fluorescent protein gives more reliable fluorescence
CC   resonance energy transfer results and are useful to detect molecule
CC   interaction, enzymes, or sample pH. These are also used to identify
CC   agents or conditions that regulate expression of control sequences. The
CC   present sequence is Discosoma sp. red fluorescent protein (RFP) mutant.
CC   Note: This sequence is not shown in the specification, however this
CC   sequence is constructed based on Discosoma sp. wild-type RFP protein
CC   shown in page 30-31 (ABW00918)
XX
SQ   Sequence 225 AA;

Qy   Query Match   99.3%;  Score 1206;  DB 7;  Length 225;
Best Local Similarity   99.6%;  Pred. No. 8.2e-127;
Matches 224;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 51
ID   ADH34501 standard; protein; 225 AA.
XX
AC   ADH34501;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Discosoma sp. DsRed mutant S148A/K167M.
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   fluorescence activated cell sorting; fluorescent timer;
KW   red fluorescent protein; DsRed; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  146
FT                    /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT                    residue 148"
FT   Misc-difference  163
FT                    /note= "Met replaces wild-type Lys. Corresponds to GFP
FT                    residue 167"
XX
PN   WO200305T833-A2.
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.
XX
PR   26-DEC-2001; 2001US-0343128P.
```

XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensor, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX

CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably a non-bioluminescent
CC   Cnidarian species, and most preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants are useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as subscreens and selective filters. Chromoprotein
CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g. transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g. green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence
CC   represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC   in an example of the invention. The present sequence is not shown in the
CC   specification, but was derived from the wild-type DsRed sequence
CC   (ADH34489) shown in Fig 1 and the information provided on page 42.
XX

SQ   Sequence 225 AA;

      Query Match   99.3%; Score 1205; DB 7; Length 225;
      Best Local Similarity   99.1%; Pred. No.1.1e-126;
      Matches 223; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEAATERLYPRDGVLKGEIHMALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
             |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 52
ABP56678
ID   ABP56678 standard; protein; 240 AA.
XX
AC   ABP56678;
XX
DT   25-MAR-2003 (first entry)
XX
DE   Mammalian codon optimised Discosoma red fluorescent protein.
XX
KW   Mammalian codon optimised Discosoma red fluorescent protein; Discosoma;
KW   red fluorescent protein; directed evolution.
XX
OS   Discosoma sp.
OS   Mammalia.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 24
FT   /note= "possible mutations at this point are N24S, N24R
FT   and N24H"
FT   Misc-difference 125
FT   /note= "possible mutations at this point are F125L and
FT   F125V"
FT   Misc-difference 164
FT   /note= "a possible mutation at this point is K164M"
FT   Misc-difference 183
FT   /note= "a possible mutation at this point is M183K"
FT   Misc-difference 206
FT   /note= "encoded by AAC"
XX
PN   WO20029492-A2.
XX
PD   28-NOV-2002.
XX
PF   20-MAY-2002; 2002WO-US015968.
XX
PR   18-MAY-2001; 2001US-0291871P.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Peelle B;
XX
DR   WPI; 2003-120798/11.
DR   N-PSDB; AB224476.
XX
PT   New Discosoma red fluorescent protein, useful for functional screens as a
PT   reporter for gene transcription, for target characterization and
PT   localization of fusion proteins, or for scaffolds for protein and peptide
PT   libraries.
XX
PS   Claim 1; Fig 1; 22pp; English.
XX

CC   The present invention describes an isolated Discosoma red fluorescent
CC   protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC   SI), with one or more point mutations at amino acid position N24, F125,
CC   K164, or M183. Also described: (1) a fusion protein comprising (1); (2)
CC   an isolated nucleic acid encoding (I); (3) a vector comprising the
CC   nucleic acid of (2); (4) a host cell comprising the nucleic acid of
CC   retroviral cDNA expression library comprising the nucleic acid of (2), or
CC   encoding (I); and (6) methods of making a fluorescent variant. The
CC   Discosoma red fluorescent proteins are useful for functional screens as a
CC   reporter for gene transcription (e.g. as a fusion protein), for target
CC   characterisation and localisation of fusion proteins, or for scaffolds
CC   for protein and peptide libraries. The fluorescent proteins can also be
CC   used as selectable markers or reporter molecules for a variety of
CC   bioassays, including methods that use fluorescence activated cell sorting
CC   (FACS) as a selection mechanism.

CC   is useful for obtaining improved variants of red fluorescent protein. The
CC   variants of Discosoma red fluorescent protein have greatly improved
CC   brightness, expression, and/or folding kinetics as compared to wild type
CC   or a codon optimised variant. The present sequence represents a mammalian
CC   codon optimised Discosoma red fluorescent protein, which is used in an
CC   example from the present invention
XX
SQ   Sequence 240 AA;

Query Match   99.3%; Score 1205; DB 6; Length 240;
Best Local Similarity 99.1%; Pred. No. 1.2e-126;
Matches 223; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIBGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||||||||||||||||||||||||||
Db   182 YMAKKPVQLPGYYYVDSKLDITSHHEDYTIVEQYERTEGRHHLFL 226

RESULT 53
AC   AAE28923;
ID   AAE28923 standard; protein; 225 AA.
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, E83 (N42H+V71A+I180H).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescent activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 42
FT   /note= "Wild-type Asn substituted with His"
FT   Misc-difference 71
FT   /note= "Wild-type Val substituted with Ala"
FT   Misc-difference 180
FT   /note= "Wild-type Ile substituted with Val"
XX
PN   WO20020268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-0006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Yanushevich Y, Savitsky A, Fradkov A;
PI   Lukyanov K,
XX
DR   WPI; 2002-691654/74.
XX

PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. or as fluorescent timers. The present
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   99.2%; Score 1204; DB 5; Length 225;
Best Local Similarity 98.7%; Pred. No. 1.4e-126;
Matches 222; Conservative   2; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKAYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 54
AC   AAE28925;
ID   AAE28925 standard; protein; 225 AA.
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, AG4 (V71M+V105A+S197T).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 71
FT   /note= "Wild-type Val substituted with Met"
FT   Misc-difference 105
FT   /note= "Wild-type Val substituted with Ala"
FT   Misc-difference 197
FT   /note= "Wild-type Ser substituted with Thr"
XX
PN   WO20020268459-A2.
XX

PD   06-SEP-2002.
XX
PP   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PA   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
XX
PN   WO020268459-A2.
XX
PT   New nucleic acid encoding a non-aggregating Cnidarian chromo- or fluorescent mutant
PT   of an aggregating chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo-/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in protease cleavage assays or as fluorescent timers.
CC   The present
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   99.2%; Score 1204; DB 5; Length 225;
Best Local Similarity   98.7%; Pred. No. 1.4e-126;
Matches 222; Conservative   2; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKMYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 55
AAE28924
ID   AAE28924 standard; protein; 225 AA.
XX
AC   AAE28924;
XX
DT   27-DEC-2002   (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, E57 (V105A+I161T+S197A).
KW   fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescence timer;
KW   bioseneor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
KW   mutein.

OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 105
FT   /note= "Wild-type Val substituted with Ala"
FT   Misc-difference 161
FT   /note= "Wild-type Ile substituted with Thr"
FT   Misc-difference 197
FT   /note= "Wild-type Ser substituted with Ala"
XX
DR   WO020268459-A2.
XX
PD   06-SEP-2002.
XX
PP   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PA   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in protease cleavage assays or as fluorescent
CC   timers. This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   99.0%; Score 1202; DB 5; Length 225;
Best Local Similarity   98.7%; Pred. No. 2.3e-126;
Matches 222; Conservative   1; Mismatches   2; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 56
ADH34503
ID   ADH34503 standard; protein; 225 AA.
XX

AC   ADH34503;
XX
DT   11-MAR-2004   (first entry)
XX
DE   Discosoma sp. DsRed mutant S148A/I165S/S203A.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   fluorescence activated cell sorting; fluorescent timer;
KW   red fluorescent protein; DsRed; mutant; mutein.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 146
FT   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT   residue 148"
FT   Misc-difference 161
FT   /note= "Ser replaces wild-type Ile. Corresponds to GFP
FT   residue 165"
FT   Misc-difference 197
FT   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT   residue 203"
XX
PN   WO2003057833-A2.
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.
PR   26-DEC-2001; 2001US-0343128P.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
PA   (CLON-) CLONTECH LAB INC.
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo- or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoprotein activity being able to be converted into fluorescent protein
CC   activity in a similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants are useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can be useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as sunscreens and selective filters. Chromoprotein

CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g. transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g. green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence
CC   represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC   in an example of the invention. The present sequence is not shown in the
CC   specification, but was derived from the wild-type DsRed sequence
CC   (ADH34489) shown in Fig 1 and the information provided on page 42.
XX
SQ   Sequence 225 AA;

Query Match   99.0%; Score 1202; DB 7; Length 225;
Best Local Similarity   98.7%; Pred. No. 2.3e-126;
Matches 222; Conservative   2; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||:|||||||||||||||||||||||||||:||||||||||||||||||||
Db   61 LSPQFQYGSKVVVKHPADIPDYKKLSFPEGFKWERVVNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEAATERLYPRDGVLKGESHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||:|||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 57
ADH34502
ID   ADH34502 standard; protein; 225 AA.
XX
AC   ADH34502;
XX
DT   11-MAR-2004   (first entry)
XX
DE   Discosoma sp. DsRed mutant S148A/K167M/S203A.
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   fluorescence activated cell sorting; fluorescent timer;
KW   red fluorescent protein; DsRed; mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 146
FT   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT   residue 148"
FT   Misc-difference 163
FT   /note= "Met replaces wild-type Lys. Corresponds to GFP
FT   residue 167"
FT   Misc-difference 197
FT   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT   residue 203"
XX
PN   WO2003057833-A2.

```
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.
PR   26-DEC-2001; 2001US-0343128P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME, Chudakov D, Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably a non-bioluminescent
CC   Cnidarian species, and most preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues in these key positions in, for example, a fluorescent protein,
CC   to those found in a chromoprotein is therefore proposed to confer
CC   chromoprotein activity on the fluorescent protein mutant, with
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants having fluorescent protein activity useful in any application
CC   that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as sunscreens and selective filters. Chromoprotein
CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g. transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g. green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of the
CC   activation or deactivation of gene expression. The present sequence
CC   represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC   in an example of the invention. The present sequence is not shown in the
CC   specification, but was derived from the wild-type DsRed sequence.
CC   (ADH34489) shown in Fig 1 and the information provided on page 42.
XX
SQ   Sequence 225 AA;

Qy   Query Match             99.0%;  Score 1202;  DB 7;  Length 225;
       Best Local Similarity   98.7%;  Pred. No.2.3e-126;
Matches  222;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy           1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60

Qy          61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy         121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
                |||||||||||||||||||||||||| :|||||||||||||| ||||||||||||||||
Db         121  KVKFIGVNFPSDGPVMQKKTMGWEAATERLYPRDGVLKGEIHMALKLKDGGHYLVEFKSI  180

Qy         181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
                |||||||||||||||| ||||||||||||||||||||||||||||
Db         181  YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL  225

RESULT 58
ID   ABP56685 standard; protein; 240 AA.
XX
AC   ABP56685;
XX
DT   25-MAR-2003  (first entry)
XX
DE   Discosoma red fluorescent protein variant K93R.
XX
KW   Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW   Discosoma; red fluorescent protein; directed evolution; mutant.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
PN   WO2002294992-A2.
XX
PD   28-NOV-2002.
XX
PF   20-MAY-2002; 2002WO-US015968.
PR   18-MAY-2001; 2001US-0291871P.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Peelle B;
XX
DR   WPI; 2003-120798/11.
XX
PT   New Discosoma red fluorescent protein, useful for functional screens as a
PT   reporter for gene transcription, for target characterisation and
PT   localization of fusion proteins, or for scaffolds for protein and peptide
PT   libraries.
XX
PS   Claim 17; Page; 22pp; English.
XX
CC   The present invention describes an isolated Discosoma red fluorescent
CC   protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC   S1), with one or more point mutations at amino acid position N24, F125,
CC   K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC   an isolated nucleic acid encoding (I); (3) a vector comprising the
CC   nucleic acid of (2); (4) a host cell comprising the vector; (5) a
CC   retroviral cDNA expression library comprising the nucleic acid of (2), or
CC   encoding (I); and (6) methods of making a fluorescent variant. The
CC   Discosoma red fluorescent proteins are useful for functional variant. The
CC   reporter for gene transcription (e.g. as a fusion protein), for target
CC   characterisation and localisation of fusion proteins, or for scaffolds
CC   for protein and peptide libraries. The fluorescent proteins can also be
CC   used as selectable markers or reporter molecules for a variety of
CC   bioassays, including methods that use fluorescence activated cell sorting
CC   (FACS) as a selection mechanism. The method of directed protein evolution
CC   is useful for obtaining improved variants of red fluorescent protein. The
CC   variants of Discosoma red fluorescent protein have greatly improved
CC   brightness, expression, and/or folding kinetics as compared to wild type
CC   or a codon optimised variant. The present sequence represents a mammalian
CC   codon optimised Discosoma red fluorescent protein variant from the
CC   present invention. N.B. The present sequence is not given in the
CC   specification, but is derived from the protein in Fig 1 as stated in the
CC   claims.
XX
SQ   Sequence 240 AA;
```

Query Match 99.0%; Score 1202; DB 6; Length 240;
Best Local Similarity 98.7%; Pred. No. 2.5e-126;
Matches 222; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFRWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db  182 YMAKKPVQLPGYYYVDSKLDITSHHEDYTIVBQYERTEGRHHLFL 226
```

RESULT 59
ABP70039

ID   ABP70039 standard; protein; 225 AA.
XX
AC   ABP70039;
XX
DT   06-AUG-2003   (revised)
DT   22-JAN-2003   (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 241.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO20270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-02732227P.
PR   21-MAR-2001; 2001AU-0003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI   Hoegh-Guldberg IO, Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts an altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Example 19; Page 499; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant

CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience phototropic UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 225 AA;

Query Match 98.8%; Score 1200; DB 5; Length 225;
Best Local Similarity 98.7%; Pred. No. 3.9e-126;
Matches 222; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGNKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||:|||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 60
AAB28926

ID   AAB28926 standard; protein; 225 AA.
XX
AC   AAB28926;
XX
DT   27-DEC-2002   (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant protein, (V71M+V105A+Y120H+S197T).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
KW   mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 71
FT   /note= "Wild-type Val substituted with Met"
FT   Misc-difference 105
FT   /note= "Wild-type Val substituted with Ala"
FT   Misc-difference 120
FT   /note= "Wild-type Tyr substituted with His"
FT   Misc-difference 197
FT   /note= "Wild-type Ser substituted with Thr"
XX
PN   WO200268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
```

XX
PR 21-FEB-2001; 2001US-0270983P.
PR 04-DEC-2001; 2001US-00006922.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
FT WPI; 2002-691654/74.
XX
DR WPI; 2002-691654/74.
XX
PT New nucleic acid encoding a non-aggregating Cnidarian chromo- or
PT fluorescent protein or mutant for
PT of an aggregating Cnidarian chromo- or fluorescent protein or mutant
PT analyte detection assays or fluorescence activated cell sorting
PT applications.
XX
PS Disclosure; Page; 80pp; English.
XX
CC The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays, as sunscreens or filters, as markers in
CC recombinant DNA applications, as colouring agents, as markers in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC invention is Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC invention. Note: This sequence is not shown in the specification, but is
CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ Sequence 225 AA;
XX
Query Match 98.8%; Score 1199; DB 5; Length 225;
Best Local Similarity 98.2%; Pred. No. 5e-126;
Matches 221; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
        ||||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVQDSSLQDGCFIY  120
        |||||||||:|||||||||||||||||||||||||||||||| |:|||||||||||||:
Db   61 LSPQFQYGSKMYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIH  120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
        |||||||||||||||| |||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

RESULT 61

ADH34505

ID ADH34505 standard; protein; 225 AA.
XX
AC ADH34505;
XX
DT 11-MAR-2004 (first entry)
XX
DE Discosoma sp. DsRed mutant S148C/I165N/S203A.
XX
KW Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW Cnidarian; Anthozoan; labelling; colouring agent; pigment;
KW analyte detection assay; selectable marker; sunscreen; selective filter;
KW fluorescence resonance energy transfer; FRET; biosensor;
KW whole cell marker; second messenger detector; in vivo marker;
KW fluorescence activated cell sorting; fluorescent timer;
KW red fluorescent protein; DsRed; mutant; mutein.
OS Synthetic.
OS Discosoma sp.

XX
Key Location/Qualifiers
FH
FT Misc-difference 146
FT /note= "Cys replaces wild-type Ser. Corresponds to GFP
FT residue 148"
FT Misc-difference 161
FT /note= "Asn replaces wild-type Ile. Corresponds to GFP
FT residue 165"
FT Misc-difference 197
FT /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT residue 203"
XX
PN WO200305783-A2.
XX
PD 17-JUL-2003.
XX
PF 23-DEC-2002; 2002WO-US041418.
PR 26-DEC-2001; 2001US-0343128P.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Bulina ME, Chudakov D, Lukyanov KA;
XX
DR WPI; 2003-607998/57.
XX
PT Novel nucleic acid encoding interconverted mutant of chromo- or
PT fluorescent protein which are useful as biosensors, coloring agents.
XX
PS Example 1; Page; 56pp; English.
XX
CC The invention relates to interconverted mutants of chromoproteins (CP) or
CC fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC derived from a Cnidarian species, preferably a non-bioluminescent
CC Cnidarian species, and most preferably an Anthozoan species. The
CC invention is based on the finding that although green fluorescent protein
CC (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC homology, there are certain positions (referred to as 148, 165, 167 and
CC 203; numbering corresponds to GFP) that are occupied by noticeably
CC different residues in the two types of proteins. Mutagenesis of the
CC residues in these key positions in, for example, a fluorescent protein,
CC to those found in a chromoprotein is therefore proposed to confer
CC chromoprotein activity on the fluorescent protein mutant, with
CC chromoproteins being able to be converted into fluorescent proteins in a
CC similar manner. The invention also relates to expression constructs,
CC vectors, host cells and host cell progeny comprising a nucleic acid of
CC the invention; the recombinant production of an interconverted
CC chromoprotein or fluorescent protein mutant; and antibodies specific for
CC interconverted mutant proteins of the invention. The interconverted
CC mutants are useful in any application that employs a chromoprotein or
CC fluorescent protein. Fluorescent protein mutants having chromoprotein or
CC chromoprotein activity can be useful as colouring agents in, for example, food
CC compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC with chromoprotein activity are also useful as labels in biological
CC analyte detection assays, as selectable markers in recombinant DNA
CC applications (e.g. the production of transgenic cells and organisms), and
CC are also useful as sunscreens and selective filters. Chromoprotein
CC mutants having fluorescent protein activity are also useful in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC changes in multicellular reorganisation and migration, as second
CC messenger detectors, as in vivo markers in animals (e.g. transgenic
CC animals), in fluorescence activated cell sorting applications, in
CC protease cleavage assays, and in assays to determine the phospholipid
CC composition in biological membranes. Proteins with fluorescent protein
CC activity can also be used as fluorescent timers, where the switch of one
CC fluorescent colour to another (e.g., green to red) is concomitant with
CC the ageing of the protein and is useful for determination of the
CC activation or deactivation of gene expression. The present sequence
CC represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC in an example of the invention. The present sequence is not shown in the
CC specification, but was derived from the wild-type DsRed sequence
CC (ADH34489) shown in Fig 1 and the information provided on page 42.

```
XX
SQ  Sequence 225 AA;

Query Match        98.8%;  Score 1199;  DB 7;  Length 225;
Best Local Similarity 98.7%;  Pred. No. 5e-126;
Matches 222;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 62
ID  ABP56684 standard; protein; 240 AA.
AC  ABP56684;
XX
DT  25-MAR-2003  (first entry)
XX
DE  Discosoma red fluorescent protein variant F125L.
XX
KW  Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW  Discosoma; red fluorescent protein; directed evolution; mutant.
XX
OS  Discosoma sp.
OS  Synthetic.
XX
PN  WO200294992-A2.
XX
PD  28-NOV-2002.
XX
PF  20-MAY-2002; 2002WO-US015968.
XX
PR  18-MAY-2001; 2001US-0291871P.
XX
PA  (RIGE-) RIGEL PHARM INC.
PI  Peelle B;
XX
DR  WPI; 2003-120798/11.
XX
PT  New Discosoma red fluorescent protein, useful for functional screens as a
PT  reporter for gene transcription, for target characterization and
PT  localization of fusion proteins, or for scaffolds for protein and peptide
PT  libraries.
XX
PS  Claim 15; Page; 22pp; English.
XX
CC  The present invention describes an isolated Discosoma red fluorescent
CC  protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC  S1), with one or more point mutations at amino acid position N24, F125,
CC  K164, or M183. Also described: (1) a fusion protein comprising (I); (3)
CC  an isolated nucleic acid encoding (I); (3) a vector comprising the
CC  nucleic acid of (2); (4) a host cell comprising the vector; (5) a
CC  retroviral cDNA expression library comprising the nucleic acid of (2), or
CC  encoding (I); and (6) methods of making a fluorescent variant. The
CC  Discosoma red fluorescent proteins are useful for functional screens as a
CC  reporter for gene transcription (e.g. as a fusion protein), for target
CC  characterisation and localization of fusion proteins, or for scaffolds
CC  for protein and peptide libraries. The fluorescent proteins can also be
```

```
CC  used as selectable markers or reporter molecules for a variety of
CC  bioassays, including methods that use fluorescence activated cell sorting
CC  (FACS) as a selection mechanism. The method of directed protein evolution
CC  is useful for obtaining improved variants of red fluorescent protein. The
CC  variants of Discosoma red fluorescent protein have greatly improved
CC  brightness, expression, and/or folding kinetics as compared to wild type
CC  or a codon optimised variant. The present sequence represents a mammalian
CC  codon optimised Discosoma red fluorescent protein variant from the
CC  present invention. N.B. The present sequence is not given in the
CC  specification, but is derived from the protein in Fig 1 as stated in the
CC  claims
XX
SQ  Sequence 240 AA;

Query Match        98.8%;  Score 1199;  DB 6;  Length 240;
Best Local Similarity 98.7%;  Pred. No. 5e-126;
Matches 222;  Conservative 2;  Mismatches 1;  Indels 0;  Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKLIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db 182 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 63
ID  ABP56683 standard; protein; 240 AA.
AC  ABP56683;
XX
DT  25-MAR-2003  (first entry)
XX
DE  Discosoma red fluorescent protein variant K164M.
XX
KW  Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW  Discosoma; red fluorescent protein; directed evolution; mutant.
XX
OS  Discosoma sp.
OS  Synthetic.
XX
PN  WO200294992-A2.
XX
PD  28-NOV-2002.
XX
PF  20-MAY-2002; 2002WO-US015968.
XX
PR  18-MAY-2001; 2001US-0291871P.
XX
PA  (RIGE-) RIGEL PHARM INC.
PI  Peelle B;
XX
DR  WPI; 2003-120798/11.
XX
PT  New Discosoma red fluorescent protein, useful for functional screens as a
PT  reporter for gene transcription, for target characterization and
PT  localization of fusion proteins, or for scaffolds for protein and peptide
PT  libraries.
XX
PS  Claim 14; Page; 22pp; English.
XX
CC  The present invention describes an isolated Discosoma red fluorescent
```

```
CC   protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC   S1), with one or more point mutations at amino acid position N24, F125,
CC   K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC   an isolated nucleic acid encoding (I); (3) a vector comprising the
CC   nucleic acid of (2); (4) a host cell comprising the vector; (5) a
CC   retroviral cDNA expression library comprising the nucleic acid of (2), or
CC   encoding (I); and (6) methods of making a fluorescent variant. The
CC   Discosoma red fluorescent proteins are useful for functional screens as a
CC   reporter for gene transcription (e.g. as a fusion protein), for target
CC   characterisation and localisation of fusion proteins, or for scaffolds
CC   for protein and peptide libraries. The fluorescent proteins can also be
CC   used as selectable markers or reporter molecules for a variety of
CC   bioassays, including methods that use fluorescence activated cell sorting
CC   (FACS) as a selection mechanism. The method of directed protein evolution
CC   is useful for obtaining improved variants of red fluorescent protein. The
CC   variants of Discosoma red fluorescent protein have greatly improved
CC   brightness, expression, and/or folding kinetics as compared to wild type
CC   or a codon optimised variant. The present sequence represents a mammalian
CC   codon optimised Discosoma red fluorescent protein variant from the
CC   present invention. N.B. The present sequence is not given in Fig 1 as stated in the
CC   specification, but is derived from the protein in Fig 1 as stated in the
CC   claims
XX
SQ   Sequence 240 AA;

Qy   Query Match     98.8%;  Score 1199;  DB 6;  Length 240;
     Best Local Similarity   98.7%;  Pred. No. 5.5e-126;
     Matches 222;  Conservative   2;  Mismatches   1;  Indels   0;  Gaps   0;

Qy           1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
               :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2 VRSSKNVIKEFMRFKVRMBGTVNGHEFEIEGEGEGRPYEGHNTVLKVTKGGPLPFAWDI 61

Qy          61 LSPQFQYGSKVYVKHPADIDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHMALKLKDGGHYLVEFKSI 181

Qy         181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
               ||||||||||||||||||||||||||||:|||||||||||||||||
Db         182 YMAKKPVQLPGYYYVDSKLDITSHHEDYTIVEQYERTEGRHHLFL 226
```

```
RESULT 64
ID   ADH34504 standard; protein; 225 AA.
XX
AC   ADH34504;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Discosoma sp. DsRed mutant S148A/I165S/K167M/S203A.
XX
KW   Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW   Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW   analyte detection assay; selectable marker; sunscreen; selective filter;
KW   fluorescence resonance energy transfer; FRET; biosensor;
KW   whole cell marker; second messenger detector; in vivo marker;
KW   fluorescence activated cell sorting; fluorescence timer;
KW   red fluorescent protein; DsRed; mutant; mutein.
XX
OS   Synthetic.
OS   Discosoma sp.
XX
FH   Key             Location/Qualifiers
FH
FT   Misc-difference 146
FT                   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT                   residue 148"
FT   Misc-difference 161
FT                   /note= "Ser replaces wild-type Ile. Corresponds to GFP
```

```
FT                   residue 165"
FT   Misc-difference 163
FT                   /note= "Met replaces wild-type Lys. Corresponds to GFP
FT                   residue 167"
FT   Misc-difference 197
FT                   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT                   residue 203"
XX
PN   WO200305783-A2.
XX
PD   17-JUL-2003.
XX
PF   23-DEC-2002; 2002WO-US041418.
XX
PR   26-DEC-2001; 2001US-0343128P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Bulina ME,  Chudakov D,  Lukyanov KA;
XX
DR   WPI; 2003-607998/57.
XX
PT   Novel nucleic acid encoding interconverted mutant of chromo-or
PT   fluorescent protein which are useful as biosensors, coloring agents.
XX
PS   Example 1; Page; 56pp; English.
XX
CC   The invention relates to interconverted mutants of chromoproteins (CP) or
CC   fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC   derived from a Cnidarian species, preferably a non-bioluminescent
CC   Cnidarian species, and most preferably an Anthozoan species. The
CC   invention is based on the finding that although green fluorescent protein
CC   (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC   homology, there are certain positions (referred to as 148, 165, 167 and
CC   203; numbering corresponds to GFP) that are occupied by noticeably
CC   different residues in the two types of proteins. Mutagenesis of the
CC   residues found in a chromoprotein is therefore proposed to confer
CC   to those found in a fluorescent protein, with the
CC   chromoprotein activity on the fluorescent protein mutant, and vice versa.
CC   chromoproteins being able to be converted into fluorescent proteins in a
CC   similar manner. The invention also relates to expression constructs,
CC   vectors, host cells and host cell progeny comprising a nucleic acid of
CC   the invention; the recombinant production of an interconverted
CC   chromoprotein or fluorescent protein mutant; and antibodies specific for
CC   interconverted mutant proteins of the invention. The interconverted
CC   mutants are useful in any application that employs a chromoprotein or
CC   fluorescent protein. Fluorescent protein mutants having chromoprotein
CC   activity can useful as colouring agents in, for example, food
CC   compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC   with chromoprotein activity are also useful as labels in biological
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications (e.g. the production of transgenic cells and organisms), and
CC   are also useful as sunscreens and selective filters. Chromoprotein
CC   mutants having fluorescent protein activity useful in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC   changes in multicellular reorganisation and migration, as second
CC   messenger detectors, as in vivo markers in animals (e.g., transgenic
CC   animals), in fluorescence activated cell sorting applications, in
CC   protease cleavage assays, and in assays to determine the phospholipid
CC   composition in biological membranes. Proteins with fluorescent protein
CC   activity can also be used as fluorescent timers, where the switch of one
CC   fluorescent colour to another (e.g., green to red) is concomitant with
CC   the ageing of the protein and is useful for determination of gene
CC   activation or deactivation of gene expression. The present sequence
CC   represents a Discosoma sp. red fluorescent protein DsRed mutant generated
CC   in an example of the invention. The present sequence is not shown in the
CC   specification, but was derived from the wild-type DsRed sequence
CC   (ADH34489) shown in Fig 1 and the information provided on page 42.
XX
SQ   Sequence 225 AA;

Query Match              98.5%;  Score 1196;  DB 7;  Length 225;
```

RESULT 65
AD078068
ID   AD078068 standard; protein; 236 AA.
XX
AC   AD078068;
XX
DT   26-AUG-2004 (first entry)
XX
DE   Corallimorpharia red fluorescent protein mutant RedI.
XX
KW   fluorescent protein; red fluorescent protein; green fluorescent protein;
KW   gene expression detection; regeneration; multiple labeling system;
KW   in vivo marker; microinjection assay; taxonomic marker; colour indicator;
KW   food additive; cosmetic; Actinodiscus; Discosoma; mutein; mutant; RedI;
KW   coral-like anemone.
XX
OS   Corallimorpharia.
XX
PN   US2004110225-A1.
XX
PD   10-JUN-2004.
XX
PF   09-DEC-2002; 2002US-00314936.
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2004-467659/44.
DR   N-PSDB; AD078067.
XX
PT   New first mutant fluorescent protein having an optical property
PT   relatively different to a corresponding optical property produced by a
PT   wild-type red or green fluorescent protein, useful for detecting
PT   expression of a gene.
XX
PS   Claim 6; SEQ ID NO 2; 30pp; English.

The invention describes a first mutant fluorescent protein (I) comprising
an amino acid sequence that differs from that of a corresponding wild-
type red fluorescent protein by an amino acid substitution, where the
first mutation fluorescent protein mutant has an optical property that
differs relative to the corresponding optical property produced by the
wild-type red or green fluorescent protein. (I) is useful for detecting
expression of gene. The polynucleotide (II) encoding (I) is useful for
detecting expression of a gene which involves introducing (II) into a
cell or organism, allowing the cell to replicate, and detecting
expression of the nucleic acid by emission of fluorescent light. The
expression of the nucleic acid expression is detected in vivo or in
vitro. (II) is useful for regenerating an animal such as a zebrafish
which contains (II). (I) is useful as a marker for detecting expression
of a gene, in biochemical assays, and as reagents. (I) is also useful in
multiple labeling systems, as in vivo markers such as in mRNA
microinjection assays, coloured food additives, and
colour indicators in diagnostic kits, coloured markers for studies of
genetics, cosmetics ingredients. (I) is useful in research for up or down
regulation, to monitor promoter activity, to allow longer term monitoring
and to localise proteins. (I) has enhanced properties such as
substantially enhanced fluorescence and reduced toxicity. This is the
amino acid sequence of RedI, a mutant of the red fluorescent protein
isolated from an aquatic species believed to be either an Actinodiscus or
Discosoma species.
XX
SQ   Sequence 236 AA;

Qy   Query Match   98.5%;   Score 1196;   DB 8;   Length 236;
Best Local Similarity   98.2%;   Pred. No. 1.2e-125;
Matches 221;   Conservative 2;   Mismatches 2;   Indels 0;   Gaps 0;

Qy   1   MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI   180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   225
             ||||||||||||||| |||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL   225

RESULT 66
AD078070
ID   AD078070 standard; protein; 236 AA.
XX
AC   AD078070;
XX
DT   26-AUG-2004 (first entry)
XX
DE   Corallimorpharia red fluorescent protein mutant RedII.
XX
KW   fluorescent protein; red fluorescent protein; green fluorescent protein;
KW   gene expression detection; regeneration; multiple labeling system;
KW   in vivo marker; microinjection assay; taxonomic marker; colour indicator;
KW   food additive; cosmetic; Actinodiscus; Discosoma; mutein; mutant; RedII;
KW   coral-like anemone.
XX
OS   Corallimorpharia.
XX
PN   US2004110225-A1.
XX
PD   10-JUN-2004.
XX
PF   09-DEC-2002; 2002US-00314936.
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2004-467659/44.
DR   N-PSDB; AD078069.
XX
PT   New first mutant fluorescent protein having an optical property

Best Local Similarity   98.2%;   Pred. No. 1.1e-125;   Indels 0;   Gaps 0;
Matches 221;   Conservative 2;   Mismatches 2;   Indels 0;   Gaps 0;

Qy   1   MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY   120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI   180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   225
             ||||||||||||||| |||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL   225

```
PT  relatively different to a corresponding optical property produced by a
PT  wild-type red or green fluorescent protein, useful for detecting
PT  expression of a gene.
XX
PS  Claim 7; SEQ ID NO 4; 30pp; English.
XX
CC  The invention describes a first mutant fluorescent protein (I) comprising
CC  an amino acid sequence that differs from that of a corresponding wild-
CC  type red fluorescent protein by an amino acid substitution, where the
CC  first mutation fluorescent protein mutant has an optical property that
CC  differs relative to the corresponding optical property produced by the
CC  wild-type red or green fluorescent protein. (I) is useful for detecting
CC  expression of gene. The polynucleotide (II) encoding (I) is useful for
CC  detecting expression of a gene which involves introducing (II) into a
CC  cell or organism, allowing the cell to replicate, and detecting
CC  expression of the nucleic acid by emission of fluorescent light. The
CC  expression of the nucleic acid expression is detected in vivo or in
CC  vitro. (II) is useful for regenerating an animal such as a zebrafish
CC  which contains (II). (I) is useful as a marker for detecting expression
CC  of a gene, in biochemical assays, and as reagents. (I) is also useful in
CC  multiple labeling systems, as in vivo markers such as in mRNA
CC  microinjection assays, and as taxonomic markers for studies of genetics,
CC  colour indicators in diagnostic kits, coloured food additives, and
CC  cosmetics ingredients. (I) is useful in research for up or down
CC  regulation, to monitor promoter activity, to allow longer term monitoring
CC  and to localise proteins. (I) has enhanced properties such as
CC  substantially enhanced fluorescence and reduced toxicity. This is the
CC  amino acid sequence of RediI, a mutant of the red fluorescent protein
CC  isolated from an aquatic species believed to be either an Actinodiscus or
CC  Discosoma species.
XX
SQ  Sequence 236 AA;

Query Match             98.5%;  Score 1196;  DB 8;  Length 236;
Best Local Similarity   98.2%;  Pred. No. 1.2e-125;
Matches  221;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            ||||||||||||||||||||||||||||:||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225
```

```
RESULT 67
AD284223
ID   AD284223 standard; protein; 236 AA.
XX
AC   AD284223;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Actinodiscus/Discosoma RFP mutant Red I, SEQ ID NO:2.
XX
KW   Red fluorescent protein; Red I; mutein.
XX
OS   Corallimorpharia.
OS   Synthetic.
XX
FH   Key      Location/Qualifiers
FT   Misc-difference  232
FT                    /note="This residue is Asn in the mutant Ac/DsRFP Red
FT                    II"
```

```
XX
PN   US2005100954-A1.
XX
PD   12-MAY-2005.
XX
PF   23-DEC-2004; 2004US-00021014.
XX
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2005-345399/35.
DR   N-PSDB; AD284222.
XX
PS   Claim 6; SEQ ID NO 2; 29pp; English.
XX
CC   Novel mutant fluorescent protein comprising amino acid sequence that
CC   differs from wild-type red or green fluorescent protein by amino acid
CC   substitutions, useful as markers for detecting desired gene expression.
CC   The invention relates to mutant red and green fluorescent proteins which
CC   have higher fluorescent intensities compared to the wild-type proteins
CC   from which they are derived. The invention also discloses nucleic acids
CC   encoding the mutant fluorescent proteins, and transgenic animals
CC   comprising such a nucleic acid. The mutant red fluorescent proteins (RFP)
CC   of the invention, Red I (AD284223) and Red II (AD284225), are derived
CC   from a novel wild-type RFP (referred to as Ac/DsRFP) isolated from a
CC   mushroom coral believed to be either an Actinodiscus or Discosoma
CC   species, while the mutant green fluorescent proteins (GFP), Green I
CC   (AD284227) and Green (AD284229), are derived from a novel wild-type GFP
CC   (known as McGFP) isolated from the great star coral Montastraea
CC   cavernosa. The Ac/DsRFP and McGFP mutants were obtained by subjecting the
CC   wild-type Ac/DsRFP and McGFP cDNAs to low-stringency PCR to introduce
CC   random mutations, cloning the transformed bacteria having higher relative
CC   fluorescence compared to those expressing the wild-type proteins. The
CC   mutant fluorescence proteins of the invention may be used in a wide range
CC   of applications. They are useful as markers for detecting the expression
CC   of a gene of interest, or can be used as in vivo markers in mRNA
CC   microinjection assays in transgenic animals. They can be used in
CC   expression systems (e.g., as components of fusion proteins), in multiple
CC   labeling systems, or can be used as taxonomic markers for studies of
CC   cnidarian genetics, as color indicators in diagnostic kits, as colored
CC   food additives and as cosmetic ingredients. The present sequence
CC   represents the mutant Actinodiscus/Discosoma red fluorescent protein Red
CC   I.
XX
SQ   Sequence 236 AA;

Query Match             98.5%;  Score 1196;  DB 9;  Length 236;
Best Local Similarity   98.2%;  Pred. No. 1.2e-125;
Matches  221;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            ||||||||||||||||||||||||||||:||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225
```

RESULT 68

ID   AD284225 standard; protein; 236 AA.
XX
AC   AD284225;
XX
DT   14-JUL-2005 (first entry)
XX
DE   Actinodiscus/Discosoma RFP mutant Red II, SEQ ID NO:4.
XX
KW   Red fluorescent protein; Red II; mutein.
OS   Synthetic.
OS   Corallimorpharia.
XX
FH   Key   Location/Qualifiers
FT
FT   Misc-difference 232
FT   /note= "This residue is Asp in the mutant Ac/DsRFP Red I"
XX
PN   US2005100954-A1.
XX
PD   12-MAY-2005.
XX
PF   23-DEC-2004; 2004US-00021014.
XX
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2005-345399/35.
DR   N-PSDB; AD284224.
XX
PT   Novel mutant fluorescent protein comprising amino acid sequence that
PT   differs from wild-type red or green fluorescent protein by amino acid
PT   substitutions, useful as markers for detecting desired gene expression.
XX
PS   Claim 7; SEQ ID NO 4; 29pp; English.
XX
CC   The invention relates to mutant fluorescent proteins which
CC   have higher fluorescent intensities compared to the wild-type proteins
CC   from which they are derived. The invention also discloses nucleic acids
CC   encoding the mutant fluorescent proteins, and transgenic animals
CC   comprising such a nucleic acid. The mutant red fluorescent proteins (RFP)
CC   of the invention, Red I (AD284223) and Red II (AD284225), are derived
CC   from a novel wild-type RFP (referred to as Ac/DsRFP) isolated from a
CC   mushroom coral believed to be either an Actinodiscus or Discosoma
CC   species, while the mutant green fluorescent proteins (GFP), Green I
CC   (AD284227) and Green (AD284229), are derived from a novel wild-type GFP
CC   (known as McGFP) isolated from the great star coral Montastraea
CC   cavernosa. The Ac/DsRFP and McGFP mutants were obtained by subjecting the
CC   wild-type Ac/DsRFP and McGFP cDNAs to low-stringency PCR to introduce
CC   random mutations, cloning the PCR products into bacterial expression
CC   vectors, and selecting the transformed bacteria having higher relative
CC   fluorescence compared to those expressing the wild-type proteins. The
CC   mutant fluorescent proteins of the invention may be used in a wide range
CC   of applications. They are useful as markers for detecting the expression
CC   of a gene of interest, or can be used as in vivo markers in mRNA
CC   microinjection assays in transgenic animals. They can be used in
CC   expression systems (e.g., as components of fusion proteins), in multiple
CC   labeling systems, or can be used in biochemical assays and as reagents.
CC   They may additionally be used as taxonomic markers for studies of
CC   cnidarian genetics, as color indicators in diagnostic kits, as colored
CC   food additives and as cosmetic ingredients. The present sequence
CC   represents the mutant Actinodiscus/Discosoma red fluorescent protein Red
CC   II. This protein was found to have a fluorescent intensity at least 50%
CC   greater than the mutant RFP Red I (AD284223).

SQ   Sequence 236 AA;

Qy   Query Match   98.5%; Score 1196; DB 9; Length 236;
      Best Local Similarity   98.2%; Pred. No. 1.2e-125;
      Matches 221; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||::||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKXI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIIVEQYERTCGRHHLFL 225

RESULT 69

ID   ADQ59562 standard; protein; 284 AA.
XX
AC   ADQ59562;
XX
DT   07-OCT-2004 (first entry)
XX
DE   Discosoma fluorescent protein drFP583.
XX
KW   optically detectable signal; protein fragment complementation assay; PCA;
KW   molecular interaction detection; biomolecular interaction detection;
KW   fluorescent protein; multi-colour PCA; drug discovery; target validation;
KW   high-throughput screening; high-content screening; pathway mapping;
KW   drug mechanism-of-action study; biosensor; diagnostic;
KW   fluorescent protein; spectral shift.
XX
OS   Discosoma.
XX
PN   US2004137528-A1.
XX
PD   15-JUL-2004.
XX
PF   01-DEC-2003; 2003US-00724178.
XX
PR   02-FEB-1998;   98US-00017412.
PR   07-FEB-2000; 2000US-00499464.
PR   24-MAY-2002; 2002US-00154758.
PR   29-JAN-2003; 2003US-00533090.
PR   09-APR-2003; 2003US-0461133P.
XX
PA   (MICH/) WATSON MICHNICK S W.
PA   (MACD/) MACDONALD M L.
PA   (LAME/) LAMERDIN J.
XX
PI   Watson Michnick SW, Macdonald ML, Lamerdin J;
XX
DR   WPI; 2004-533363/51.
XX
PT   Composition useful in protein fragment complementation assays for drug
PT   discovery and high-throughput screening, comprising complementary
PT   fragments of protein or mutant protein, generating optically detectable
PT   signal when associated.
XX
PS   Disclosure; SEQ ID NO 12; 34pp; English.
XX
CC   The invention describes a composition (I) comprising complementary
CC   fragments of a protein or mutant protein, where the fragments generate an
CC   optically detectable signal when associated, and each of the mutant
CC   protein fragments is fused to a separate molecule. Also described are:

CC protein fragment complementation assays (PCAs) for detection of molecular
CC interaction, involving reassembling separate fragments from an optically
CC detectable protein, and detecting the reassembly by units of
CC reconstitution of activity of the optically detectable protein, where the
CC reassembly of the fragments is operated by the interaction of other molecular
CC domains fused to each fragment, and is independent of other molecular
CC processes; detecting biomolecular interaction, involving selecting an
CC appropriate optically detectable protein, effecting fragmentation of the
CC optically detectable protein such that the fragmentation results in
CC reversible loss of protein function, fusing or attaching fragments of the
CC optically detectable protein separately to other molecules, reassociating
CC the protein fragments through interactions of the molecules that are
CC fused or attached to the fragments, and detecting the resulting optical
CC signal; designing and engineering of PCAs based on fluorescent protein;
CC and a method and composition for the construction of multi-color PCAs.
CC (I) is useful in PCAs and other assays for drug discovery, target
CC validation, high-throughput screening, high-content screening, pathway
CC mapping, drug mechanism-of-action studies, biosensors and diagnostics.
CC (I) is useful for engineering different colour PCAs for a variety of
CC applications in biology and biotechnology. This is the amino acid
CC sequence of a Discosoma fluorescent protein used in a comparison of
CC fluorescent proteins with Aequorea green fluorescent protein, variants of
CC which can be created with altered excitation and emission wavelength
CC relative to wild type GFP.
XX
SQ Sequence 284 AA;

Query Match   98.5%; Score 1195.5; DB 8; Length 284;
Best Local Similarity 99.1%; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Matches 224; Pred. No. 1.8e-125;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 118

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 178

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 179 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 238

Qy 181 YMAKK-PVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         ||||| ||||||||||||||||||||||||||||||||||||||
Db 239 YMAKKAPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 284

RESULT 70
ABP56681
ID ABP56681 standard; protein; 240 AA.
XX
AC ABP56681;
XX
DT 25-MAR-2003 (first entry)
XX
DE Discosoma red fluorescent protein variant F125L and N24S.
XX
KW Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW Discosoma; red fluorescent protein; directed evolution; mutant.
XX
OS Discosoma sp.
OS Synthetic.
XX
PN WO200294992-A2.
PD 28-NOV-2002.
XX
PF 20-MAY-2002; 2002WO-US015968.
XX
PR 18-MAY-2001; 2001US-0291871P.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Peelle B;
XX
DR WPI; 2003-120798/11.
XX

New Discosoma red fluorescent protein, useful for functional screens as a
reporter for gene transcription, for target characterization and
localization of fusion proteins, or for scaffolds for protein and peptide
libraries.

Claim 11; Page; 22pp; English.

XX
SQ Sequence 240 AA;

CC The present invention describes an isolated Discosoma red fluorescent
CC protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC S1), with one or more point mutations at amino acid position N24, F125,
CC K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC an isolated nucleic acid encoding (I); (3) a vector comprising the
CC nucleic acid of (2); (4) a host cell comprising the nucleic acid of (2), or
CC retroviral cDNA expression library comprising a fluorescent variant. The
CC encoding (II); and (6) methods of making a fluorescent variant. The
CC Discosoma red fluorescent proteins are useful for functional variant. The
CC reporter for gene transcription (e.g. as a fusion protein), for target
CC characterisation and localisation of fusion proteins, or for scaffolds
CC for protein and peptide libraries. The fluorescent proteins can also be
CC used as selectable markers or reporter molecules for a variety of
CC bioassays, including methods that use fluorescence activated cell sorting
CC (FACS) as a selection mechanism. The method of directed protein evolution
CC is useful for obtaining improved variants of red fluorescent protein. The
CC variants of Discosoma red fluorescent protein have greatly improved
CC brightness, expression, and/or folding kinetics as compared to wild type
CC or a codon optimised variant. The present sequence represents a mammalian
CC codon optimised Discosoma red fluorescent protein variant from the
CC present invention. N.B. The present sequence is not given in the
CC specification, but is derived from the protein in Fig 1 as stated in the
CC claims
XX

Query Match   98.4%; Score 1194; DB 6; Length 240;
Best Local Similarity 98.2%; Pred. No. 2e-125;
Matches 221; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         :|||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVSGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKLIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 182 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 71
ADH34506
ID ADH34506 standard; protein; 225 AA.
XX
AC ADH34506;
XX
DT 11-MAR-2004 (first entry)
XX
DE Discosoma sp. DsRed mutant DsRed-NF S148C/I165N/K167M/S203A.
XX
KW Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW Cnidarian; Anthozoan; labelling; colouring agents; pigment;
KW analyte detection assay; selectable marker; sunscreen; selective filter;
KW fluorescence resonance energy transfer; FRET; biosensor;

```
KW  whole cell marker; second messenger detector; in vivo marker;
KW  fluorescence activated cell sorting; fluorescent timer;
KW  red fluorescent protein; DsRed; non-fluorescent mutant; mutant;
KW  mutein.
XX
OS  Synthetic.
OS  Discosoma sp.
XX
FH  Key              Location/Qualifiers
FH
FT  Misc-difference  146
FT                   /note= "Cys replaces wild-type Ser. Corresponds to GFP
FT                   residue 148"
FT  Misc-difference  161
FT                   /note= "Asn replaces wild-type Ile. Corresponds to GFP
FT                   residue 165"
FT  Misc-difference  163
FT                   /note= "Met replaces wild-type Lys. Corresponds to GFP
FT                   residue 167"
FT  Misc-difference  197
FT                   /note= "Ala replaces wild-type Ser. Corresponds to GFP
FT                   residue 203"
XX
PN  WO2003057833-A2.
XX
PD  17-JUL-2003.
XX
PF  23-DEC-2002; 2002WO-US041418.
XX
PR  26-DEC-2001; 2001US-0343128P.
XX
PA  (CLON-) CLONTECH LAB INC.
XX
PI  Bulina ME, Chudakov D, Lukyanov KA;
XX
DR  WPI; 2003-607998/57.
XX
PT  Novel nucleic acid encoding interconverted mutant of chromo-or
PT  fluorescent protein which are useful as biosensors, coloring agents.
XX
PS  Example 1; Page; 56pp; English.
XX
CC  The invention relates to interconverted mutants of chromoproteins (CP) or
CC  fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC  derived from a Cnidarian species, preferably an Anthozoan species.
CC  Cnidarian species, and most preferably green fluorescent protein. The
CC  invention is based on the finding that although green fluorescent protein
CC  (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC  homology, there are certain positions (referred to as 148, 165, 167 and
CC  203; numbering corresponds to GFP) that are occupied by noticeably
CC  different residues in the two types of proteins. Mutagenesis of the
CC  residues in these key positions in, for example, a fluorescent protein,
CC  to those found in a chromoprotein is therefore proposed to confer
CC  chromoprotein activity on the fluorescent protein mutant, with
CC  interconverted mutant proteins of the invention. The interconverted
CC  chromoproteins being able to be converted into fluorescent proteins in a
CC  similar manner. The invention also relates to expression constructs,
CC  vectors, host cells and host cell progeny comprising a nucleic acid of
CC  the invention; the recombinant production of an interconverted
CC  chromoprotein or fluorescent protein mutant; and antibodies specific for
CC  mutants are useful in any application that employs a chromoprotein or
CC  fluorescent protein. Fluorescent protein mutants having chromoprotein
CC  activity can useful as colouring agents in, for example, food
CC  compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC  with chromoprotein activity are also useful as labels in biological
CC  analyte detection assays, as selectable markers in recombinant DNA
CC  applications (e.g. the production of transgenic cells and organisms), and
CC  are also useful as sunscreens and selective filters. Chromoprotein
CC  mutants having fluorescent protein activity useful in fluorescence
CC  resonance energy transfer (FRET) applications, as biosensors in
CC  prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC  changes in multicellular reorganisation and migration, as second
CC  messenger detectors, as in vivo markers in animals (e.g., transgenic
CC  animals), in fluorescence activated cell sorting applications, in
```

```
CC  protease cleavage assays; and in assays to determine the phospholipid
CC  composition in biological membranes. Proteins with fluorescent protein
CC  activity can also be used as fluorescent timers, where the switch of one
CC  fluorescent colour to another (e.g. green to red) is concomitant with
CC  the ageing of the protein and is useful for determination of the
CC  activation or deactivation of gene expression. The present sequence
CC  represents a non-fluorescent Discosoma sp. red fluorescent protein DsRed
CC  mutant, DsRed-NF, which has chromoprotein activity and which was
CC  generated in an example of the invention. The present sequence is not
CC  shown in the specification, but was derived from the wild-type DsRed
CC  sequence (ADH34489) shown in Fig 1 and the information provided on page
CC  42.
XX
SQ  Sequence 225 AA;

Query Match             98.3%;  Score 1193;  DB 7;  Length 225;
Best Local Similarity   98.2%;
Matches  221;  Conservative  1;  Mismatches  3;  Indels  0;  Gaps  0;

Qy          1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60

Qy         61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy        121  KVKPIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
               ||||||||||||||||||||||||||:|||||||||||||:||:|||||||||||||||||
Db        121  KVKPIGVNFPSDGPVMQKKTMGWEACTERLYPRDGVLKGENHMALKLKDGGHYLVEFKSI  180

Qy        181  YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
               |||||||||||||||:|||||||||||||||||||||||||||||
Db        181  YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEBQYERTEGRHHLFL  225

RESULT 72
ABP56679
ID  ABP56679 standard; protein; 240 AA.
XX
AC  ABP56679;
XX
DT  25-MAR-2003  (first entry)
XX
DE  Discosoma red fluorescent protein variant F125L and M183K.
XX
KW  Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW  Discosoma; red fluorescent protein; directed evolution; mutant.
XX
OS  Discosoma sp.
OS  Synthetic.
XX
CC  WO200294992-A2.
XX
PN  WO200294992-A2.
XX
PD  28-NOV-2002.
XX
PF  20-MAY-2002; 2002WO-US015968.
XX
PR  18-MAY-2001; 2001US-0291871P.
XX
PA  (RIGE-) RIGEL PHARM INC.
XX
PI  Peelle B;
XX
DR  WPI; 2003-120798/11.
XX
PT  New Discosoma red fluorescent protein, useful for functional screens as a
PT  reporter for gene transcription, for target characterization and
PT  localization of fusion proteins, or for scaffolds for protein and peptide
PT  libraries.
XX
PS  Claim 8; Page; 22pp; English.
XX
```

CC The present invention describes an isolated Discosoma red fluorescent
CC protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC S1), with one or more point mutations at amino acid position N24, F125,
CC K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC an isolated nucleic acid of (2); (4) a host cell comprising the
CC nucleic acid of (2); (4) a host cell comprising the nucleic acid of (2), or
CC retroviral cDNA expression library comprising the nucleic acid of (2), or
CC encoding (I); and (6) methods of making a fluorescent variant. The
CC Discosoma red fluorescent proteins are useful for functional screens as a
CC reporter for gene transcription (e.g. as a fusion protein), for target
CC characterisation and localisation of fusion proteins, or for scaffolds
CC for protein and peptide libraries. The fluorescent proteins can also be
CC used as selectable markers or reporter molecules for a variety of
CC bioassays, including methods that use fluorescence activated cell sorting
CC (FACS) as a selection mechanism. The method of directed protein evolution
CC is useful for obtaining improved variants of red fluorescent protein. The
CC variants of Discosoma red fluorescent protein have greatly improved
CC brightness, expression, and/or folding kinetics as compared to wild type
CC or a codon optimised variant. The present sequence represents a mammalian
CC codon optimised Discosoma red fluorescent protein variant from the
CC present invention. N.B. The present sequence is not given in the
CC specification, but is derived from the protein in Fig 1 as stated in the
CC claims

XX
SQ Sequence 240 AA;

Query Match   98.3%; Score 1193; DB 6; Length 240;
Best Local Similarity   98.2%;
Matches 221; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKLIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 182 YKAKKPVQLPGYYYVDSKLDITSHHEDYTIVEQYERTEGRHHLFL 226

RESULT 73
ABP56680
ID ABP56680 standard; protein; 240 AA.
XX
AC ABP56680;
XX
DT 25-MAR-2003 (first entry)
XX
DE Discosoma red fluorescent protein variant F125V and M183K.
XX
KW Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW Discosoma; red fluorescent protein; directed evolution; mutant.
XX
OS Discosoma sp.
OS Synthetic.
XX
PN WO200294992-A2.
XX
PD 28-NOV-2002.
XX
PF 20-MAY-2002; 2002WO-US015968.
XX
PR 18-MAY-2001; 2001US-0291871P.
XX
PA (RIGE-) RIGEL PHARM INC.
XX

PI Peelle B;
XX
DR WPI; 2003-120798/11.
XX
PT New Discosoma red fluorescent protein, useful for functional screens as a
PT reporter for gene transcription, for target characterization and
PT localization of fusion proteins, or for scaffolds for protein and peptide
PT libraries.
XX
PS Claim 9; Page; 22pp; English.
XX

CC The present invention describes an isolated Discosoma red fluorescent
CC protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC S1), with one or more point mutations at amino acid position N24, F125,
CC K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC an isolated nucleic acid encoding (I); (3) a vector comprising the
CC nucleic acid of (2); (4) a host cell comprising the nucleic acid of (2), or
CC retroviral cDNA expression library comprising the nucleic acid of (2), or
CC encoding (I); and (6) methods of making a fluorescent variant. The
CC Discosoma red fluorescent proteins are useful for functional screens as a
CC reporter for gene transcription (e.g. as a fusion protein), for target
CC characterisation and localisation of fusion proteins, or for scaffolds
CC for protein and peptide libraries. The fluorescent proteins can also be
CC used as selectable markers or reporter molecules for a variety of
CC bioassays, including methods that use fluorescence activated cell sorting
CC (FACS) as a selection mechanism. The method of directed protein evolution
CC is useful for obtaining improved variants of red fluorescent protein. The
CC variants of Discosoma red fluorescent protein have greatly improved
CC brightness, expression, and/or folding kinetics as compared to wild type
CC or a codon optimised variant. The present sequence represents a mammalian
CC codon optimised Discosoma red fluorescent protein variant from the
CC present invention. N.B. The present sequence is not given in the
CC specification, but is derived from the protein in Fig 1 as stated in the
CC claims
XX
SQ Sequence 240 AA;

Query Match   98.2%; Score 1192; DB 6; Length 240;
Best Local Similarity   98.2%;
Matches 221; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 KVKIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db 182 YKAKKPVQLPGYYYVDSKLDITSHHEDYTIVEQYERTEGRHHLFL 226

RESULT 74
AAE28837
ID AAE28837 standard; protein; 225 AA.
XX
AC AAE28837;
XX
DT 27-DEC-2002 (first entry)
XX
DE Discosoma sp. drFP583 (NFP-6) mutant protein, E5-NA.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutein;

```
KW   mutant.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FH
FT   Misc-difference  2
FT                    /note= "Wild type Arg substituted with Ala"
FT   Misc-difference  5
FT                    /note= "Wild type Lys substituted with Glu"
FT   Misc-difference  9
FT                    /note= "Wild type Lys substituted with Thr"
FT   Misc-difference  105
FT                    /note= "Wild type Val substituted with Ala"
FT   Misc-difference  197
FT                    /note= "Wild type Ser substituted with Thr"
XX
PN   WO200268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-0006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savietsky A, Fradkov A;
XX
PT   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28831) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

  Query Match          98.1%;  Score 1191;  DB 5;  Length 225;
  Best Local Similarity  97.8%;
  Matches 220;  Conservative  2;  Mismatches 3;  Indels  0;  Gaps  0;
                                 Pred. No. 4e-125;

Qy    1  MRRSKNVIKEFMRKFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         |:|:||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

```
Db   181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225

RESULT 75
AAE17542
AAE17542
XX
ID   AAE17542 standard; protein; 225 AA.
AC   AAE17542;
XX
DT   22-APR-2002  (first entry)
XX
DE   Discosoma sp. non-aggregating mutant timer protein E5NA.
XX
KW   Fluorescent timer protein; protein movement; translocation; trafficking;
KW   promoter activity; gene expression; transgenic plant; gene modification;
KW   protein age; mutant; E5NA; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FH
FT   Misc-difference  2
FT                    /note= "Wild type Arg substituted with Ala"
FT   Misc-difference  5
FT                    /note= "Wild type Lys substituted with Glu"
FT   Misc-difference  9
FT                    /note= "Wild type Lys substituted with Thr"
XX
PN   WO200196373-A2.
XX
PD   20-DEC-2001.
XX
PF   14-JUN-2001; 2001WO-US019097.
XX
PR   13-JUN-2000; 2000US-0211607P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Fradkov AF, Terskikh A;
XX
PT   Claim 5; Fig 3; 89pp; English.
XX
DR   WPI; 2002-154595/20.
DR   N-PSDB; AAD28209.
XX
CC   New fluorescent timer proteins comprising an emission spectrum that
CC   changes over time from a first wavelength to a second wavelength, useful
CC   for monitoring intracellular protein movement, translocation, trafficking
CC   or stability.
CC   The invention relates to a fluorescent timer protein having an emission
CC   spectrum that changes over time after synthesis from a first wavelength
CC   to a second wavelength. The fluorescent timer proteins are useful in
CC   monitoring the activity of a promoter, determining the age of a protein,
CC   identifying an agent that modulates the activity of a promoter and in
CC   enriching a population of cells comprising a fluorescent timer protein.
CC   The fluorescent timer proteins are also useful for assessing gene
CC   expression during development of a multicellular organism or during
CC   cellular differentiation, in response to a drug or other inducer of
CC   promoter activity, as a reporter to serve as a read-out of promoter
CC   activity, monitoring intracellular protein movement or translocation,
CC   protein trafficking, or protein stability, to investigate temporal
CC   aspects of the activity of a regulatory element, for determining cell
CC   fate during development and organ remodelling, in spatial and temporal
CC   visualisation of newly synthesised proteins and accumulated proteins, and
CC   in distinguishing between newly formed and pre-existing structures, e.g.
CC   membrane junctions and extracellular matrix components. The fluorescent
CC   timer proteins may further be used to investigations where photobleaching
CC   techniques are employed, as detectable labels, as selectable markers, as
CC   biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC   assays, and as second messenger detectors. The nucleic acids can be used
CC   to generate transgenic, non-human plants or animals or site-specific gene
CC   modifications in cell lines. The present sequence is Discosoma sp. E5NA
```

CC non-aggregating fluorescent timer protein derived from fluorescent
CC protein E5 by substituting amino acids at positions R2A, K5E and K9T. E5
CC protein is derived from humanised wild-type Anthozoa protein drFP583 by
CC substituting Val to Ala at 105 and Ser to Thr at 197
XX
SQ Sequence 225 AA;

Qy Query Match 98.1%; Score 1191; DB 5; Length 225;
    Best Local Similarity 97.8%; Pred. No. 4e-125;
    Matches 220; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy     1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         |:||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||| ||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDIKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 76

ID ABP56682 standard; protein; 240 AA.
AC ABP56682;
XX
DT 25-MAR-2003 (first entry)
XX
DE Discosoma red fluorescent protein variant F125L, N24S and M183K.
XX
KW Discosoma; red fluorescent protein; directed evolution; mutant.
KW Mammalian codon optimised Discosoma red fluorescent protein; variant;
XX
OS Discosoma sp.
OS Synthetic.
XX
PN WO20020294992-A2.
XX
PD 28-NOV-2002.
XX
PF 20-MAY-2002; 2002WO-US015968.
PR 18-MAY-2001; 2001US-0291871P.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Peelle B;
XX
DR WPI; 2003-120798/11.
XX
PT New Discosoma red fluorescent protein, useful for functional screens as a
PT reporter for gene transcription, for target characterization and
PT localization of fusion proteins, or for scaffolds for protein and peptide
PT libraries.
XX
PS Claim 13; Page; 22pp; English.
XX
CC The present invention describes an isolated Discosoma red fluorescent
CC protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC S1), with one or more point mutations at amino acid position N24, F125,
CC K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC an isolated nucleic acid encoding (I); (3) a vector comprising the
CC nucleic acid of (2); (4) a host cell comprising the vector; (5) a
CC retroviral cDNA expression library comprising the nucleic acid of (2), or
CC encoding (I); and (6) methods of making a fluorescent variant. The

---

CC Discosoma red fluorescent proteins are useful for functional screens as a
CC reporter for gene transcription (e.g. as a fusion protein), for target
CC characterisation and localisation of fusion proteins, or for scaffolds
CC for protein and peptide libraries. The fluorescent proteins can also be
CC used as selectable markers or reporter molecules for a variety of
CC bioassays, including methods that use fluorescence activated cell sorting
CC (FACS) as a selection mechanism. The method of directed protein evolution
CC is useful for obtaining improved variants of red fluorescent protein. The
CC variants of Discosoma red fluorescent protein have greatly improved
CC brightness, expression, and/or folding kinetics as compared to wild type
CC or a codon optimised variant. The present fluorescent protein variant from the
CC codon optimised Discosoma red fluorescent protein variant represents a
CC present invention. N.B. The present sequence is not given in the
CC specification, but is derived from the protein in Fig 1 as stated in the
CC claims
XX
SQ Sequence 240 AA;

Qy Query Match 97.9%; Score 1188; DB 6; Length 240;
    Best Local Similarity 97.8%; Pred. No. 9.6e-125;
    Matches 220; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy     1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         :||:||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db     2 VRSSKNVIKEFMRFKVRMEGTVSGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 121

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 KVKLIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||| ||||||||||||||||||||||||||||
Db   182 YKAKKPVQLPGYYYVDIKLDITSHNEDYTIVEQYERTEGRHHLFL 226
```

RESULT 77

ID AAE28836 standard; protein; 225 AA.
XX
AC AAE28836;
XX
DT 27-DEC-2002 (first entry)
XX
DE Discosoma sp. drFP583 (NFP-6) mutant protein, E57-NA.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filler;
KW fluorescence activated cell sorting application; fluorescence timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutein;
KW mutant.
XX
OS Discosoma sp.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FH
FT Misc-difference   2
FT   /note= "Wild type Arg substituted with Ala"
FT Misc-difference   5
FT   /note= "Wild type Lys substituted with Glu"
FT Misc-difference   9
FT   /note= "Wild type Lys substituted with Thr"
FT Misc-difference   105
FT   /note= "Wild type Val substituted with Ala"
FT Misc-difference   161
FT   /note= "Wild type Ile substituted with Thr"
FT Misc-difference   197
FT   /note= "Wild type Ser substituted with Ala"
XX

PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PD   20-FEB-2002; 2002WO-US005749.
XX
PF   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savietsky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   N-PSDB; AAD46281.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
CC   ID NO:8 (AAE28833) in page 70-71 of the specification
XX
SQ   Sequence 225 AA;

Query Match   97.7%; Score 1186; DB 5; Length 225;
Best Local Similarity   97.3%; Pred. No. 1.5e-124;
Matches 219; Conservative 2; Mismatches 4; Indels   0; Gaps   0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |:|:|| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||:|||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFPL 225
```

RESULT 78
ADH10227
ID   ADH10227 standard; protein; 225 AA.
XX
AC   ADH10227;
XX
DT   11-MAR-2004 (first entry)
XX
DE   Red fluorescent protein variant (DsRed).
XX
PF   Protein immobilisation; red fluorescent protein; DsRed.
XX
KW   Synthetic.
XX
OS   Synthetic.
XX

PN   EP1367121-A2.
XX
PD   03-DEC-2003.
XX
PF   23-MAY-2003; 2003EP-00253252.
XX
PR   23-MAY-2002; 2002JP-00148950.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
PI   Iwakura M, Hirota K;
XX
DR   WPI; 2004-045224/05.
XX
PT   New carrier, useful for immobilizing polypeptides via the carboxyl group
PT   of the C-terminus, comprises a protein immobilized adsorptively, and
PT   having a sulfhydryl group.
XX
PS   Example; SEQ ID NO 7; 31pp; English.
XX
CC   The invention relates to a carrier comprising a protein immobilized
CC   adsorptively, where the protein has a sulfhydryl group. The carrier is
CC   useful for immobilizing polypeptides via the carboxyl group of the C-
CC   terminus. The present sequence represents a Red fluorescent protein
CC   variant (DsRed) sequence.
XX
SQ   Sequence 225 AA;

Query Match   97.7%; Score 1186; DB 8; Length 225;
Best Local Similarity   97.3%; Pred. No. 1.5e-124;
Matches 219; Conservative 2; Mismatches 4; Indels   0; Gaps   0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |:|:|| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||:|||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFPL 225
```

RESULT 79
ADT77526
ID   ADT77526 standard; protein; 225 AA.
XX
AC   ADT77526;
XX
DT   13-JAN-2005 (first entry)
XX
DE   Protein SEQ ID NO:2 used in protein array.
XX
PF   Protein array; planar substrate; immobilisation; detection;
XX
KW   target substance.
XX
OS   Unidentified.
XX
PN   WO200409542-A1.
XX
PD   21-OCT-2004.
XX
PF   09-APR-2004; 2004WO-JP005150.
XX
PR   10-APR-2003; 2003JP-00106450.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

```
XX
PI   Iwakura M, Hirota K;
XX
DR   WPI; 2004-766496/75.
XX
PT   Substrate for protein array, obtained by combining a polymer compound
PT   having primary amino groups in repeating units on surface of planar
PT   substrate.
XX
PS   Example 1; SEQ ID NO 2; 32pp; Japanese.
XX
CC   The invention relates to a substrate for a protein array, obtained by
CC   combining a polymer compound having primary amino groups in repeating
CC   units on a surface of a planar substrate. The invention also relates to a
CC   protein array formed by bonding the C-termini of proteins to the primary
CC   amino groups of the polymer compound on the substrate. The invention
CC   further relates to a method for the preparation of the protein array. The
CC   preferably, the planar substrate has water absorptivity, the planar
CC   is a polyalkyl amine or a polylysine, and the immobilised proteins
CC   contain a linker peptide. The protein array of the invention enables the
CC   effective and highly sensitive detection of various target substances.
CC   Sequences ADT77525-ADT77526 represent proteins whose C-termini were
CC   joined to a negatively charged acidic linker peptide (ADT77529) to
CC   produce the sequences shown in ADT77527-ADT77528. Sequences ADT77527-
CC   ADT77528 were subsequently immobilised on a planar substrate via the
CC   linker peptide to form a protein array.
XX
SQ   Sequence 225 AA;

Query Match      97.7%;   Score 1186;  DB 8;  Length 225;
Best Local Similarity   97.3%;  Pred. No. 1.5e-124;
Matches 219;  Conservative   2;  Mismatches   4;  Indels   0;  Gaps   0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
          ||||:||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGETHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
          |||||||||||||| ||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL  225


RESULT 80
ADH10228
ID   ADH10228 standard; protein; 239 AA.
XX
AC   ADH10228;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Immobilised red fluorescent protein variant (DsRed).
XX
KW   Protein immobilisation; red fluorescent protein; DsRed.
XX
OS   Synthetic.
XX
PN   EP1367121-A2.
XX
PD   03-DEC-2003.
XX
PF   23-MAY-2003; 2003EP-00253252.
XX
PR   23-MAY-2002; 2002JP-00148950.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
```

```
XX
PI   Iwakura M, Hirota K;
XX
DR   WPI; 2004-045224/05.
XX
PT   New carrier, useful for immobilizing polypeptides via the carboxyl group
PT   of the C-terminus, comprises a protein immobilized adsorptively, and
PT   having a sulfhydryl group.
XX
PS   Example 3; SEQ ID NO 8; 31pp; English.
XX
CC   The invention relates to a carrier comprising a protein immobilized
CC   adsorptively, where the protein has a sulfhydryl group. The carrier is
CC   useful for immobilizing polypeptides via the carboxyl group of the C-
CC   terminus. The present sequence represents an immobilised red fluorescent
CC   protein variant (DsRed) sequence.
XX
SQ   Sequence 239 AA;

Query Match      97.7%;   Score 1186;  DB 8;  Length 239;
Best Local Similarity   97.3%;  Pred. No. 1.6e-124;
Matches 219;  Conservative   2;  Mismatches   4;  Indels   0;  Gaps   0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
          ||||:||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGETHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
          |||||||||||||| ||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL  225


RESULT 81
ADT77528
ID   ADT77528 standard; protein; 239 AA.
XX
AC   ADT77528;
XX
DT   13-JAN-2005  (first entry)
XX
FH   Key             Location/Qualifiers
FT   Protein         1..225
FT                   /note= "Corresponds to SEQ ID NO:2"
FT   Peptide         226..239
FT                   /note= "Corresponds to the linker peptide shown on page
FT                   12 of the specification"
XX
OS   Synthetic.
OS   Unidentified.
XX
KW   Protein array; planar substrate; immobilisation; detection;
KW   target substance; linker peptide.
XX
DE   Protein/linker peptide SEQ ID NO:4 used in protein array.
XX
PN   WO200409542-A1.
XX
PD   21-OCT-2004.
XX
PF   09-APR-2004; 2004WO-JP005150.
XX
PR   10-APR-2003; 2003JP-00106450.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
```

PI Iwakura M, Hirota K;
XX
DR WPI; 2004-766496/75.
XX
PT Substrate for protein array, obtained by combining a polymer compound
PT having primary amino groups in repeating units on surface of planar
PT substrate.
XX
PN Example 1; SEQ ID NO 4; 32pp; Japanese.
XX
PS
CC The invention relates to a substrate for a protein array, obtained by
CC combining a polymer compound having primary amino groups in repeating
CC units on a surface of a planar substrate. The invention also relates to a
CC protein array formed by bonding the C-termini of proteins to the primary
CC amino groups of the polymer compound on the substrate. The invention
CC further relates to a method for the preparation of the protein array.
CC preferably, the planar substrate has water absorptivity, the polymer used
CC is a polyallyl amine or a polylysine, and the immobilised protein
CC contain a linker peptide. The protein array of the invention enables the
CC effective and highly sensitive detection of various target substances.
CC Sequences ADT77527-ADT77528 represent proteins comprising a C-terminal
CC negatively charged acidic linker peptide (ADT77529) which were
CC immobilised on a planar substrate via the linker peptide to form a
CC protein array.
XX
SQ Sequence 239 AA;

Query Match   97.7%; Score 1186; DB 8; Length 239;
Best Local Similarity 97.7%;
Matches 219; Conservative   2; Mismatches   4; Indels   0; Gaps   0;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||| |||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy  61 LSPQFQYGSKVYVKHPADIDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||| ||||||||||||||||||||| ||||||||||||||
Db  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGETHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||| |||||||||||||||||||||||||||
Db 181 YMAKKPVQLPGYYYVVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 82
ABP56686
ID ABP56686 standard; protein; 240 AA.
XX
AC ABP56686;
XX
DT 25-MAR-2003 (first entry)
XX
DE Discosoma red fluorescent protein variant R18H,N24S,P125L and M183K.
KW Mammalian codon optimised Discosoma red fluorescent protein; variant;
KW Discosoma; red fluorescent protein; directed evolution; mutant.
OS Synthetic.
XX
OS Discosoma sp.
XX
PN WO2002094992-A2.
PD 28-NOV-2002.
XX
PF 20-MAY-2002; 2002WO-US015968.
XX
PR 18-MAY-2001; 2001US-0291871P.
XX
PA (RIGE-) RIGEL PHARM INC.

XX
PI Peelle B;
XX
DR WPI; 2003-120798/11.
XX
PT New Discosoma red fluorescent protein, useful for functional screens as a
PT reporter for gene transcription, for target characterization and
PT localization of fusion proteins, or for scaffolds for protein and peptide
PT libraries.
XX
PN Example 1; Page; 22pp; English.
XX
PS
CC The present invention describes an isolated Discosoma red fluorescent
CC protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC S1), with one or more point mutations at amino acid position N24, F125,
CC K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC an isolated nucleic acid encoding (I); (3) a vector comprising the
CC nucleic acid of (2); (4) a host cell comprising the vector; (5) a
CC retroviral cDNA expression library comprising the nucleic acid of (2), or
CC encoding (1); and (6) methods of making a fluorescent variant. The
CC Discosoma red fluorescent proteins are useful for functional screens as a
CC reporter for gene transcription (e.g. as a fusion protein), for target
CC characterisation and localisation of fusion proteins, or for scaffolds
CC for protein and peptide libraries. The fluorescent proteins can also be
CC used as selectable markers or reporter molecules for a variety of
CC bioassays, including methods that use fluorescence activated cell sorting
CC (FACS) as a selection mechanism. The method of directed protein evolution
CC is useful for obtaining improved variants of red fluorescent protein. The
CC variants of Discosoma red fluorescent protein have greatly improved
CC brightness, expression, and/or folding kinetics as compared to wild type
CC or a codon optimised variant. The present sequence represents a mammalian
CC codon optimised Discosoma red fluorescent protein variant from the
CC present invention. N.B. The present sequence is not given in the
CC specification, but is derived from the protein in Fig 1 as stated in Fig
CC 2
XX
SQ Sequence 240 AA;

Query Match   97.4%; Score 1183; DB 6; Length 240;
Best Local Similarity 97.4%;
Matches 219; Conservative   3; Mismatches   3; Indels   0; Gaps   0;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||: |||||||:||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRHMEGTVSGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 121

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        || ||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 122 KVKLIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGETHKALKLKDGGHYLVEFKSI 181

Qy 181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |:||||||||||||||| |||||||:|||||||||||||||||||
Db 182 YKAKKPVQLPGYYYVVDAKLDITSHHEDYTIVEQYERTEGRHHLFL 226
```

RESULT 83
ADL26532
ID ADL26532 standard; protein; 226 AA.
XX
AC ADL26532;
XX
DT 20-MAY-2004 (first entry)
XX
DE Discosoma fluorescent protein dsFP583.
KW fluorescent protein; fluorescent marker; gene expression;
KW protein localization; FP.
OS Discosoma sp.

```
PN   WO200401648-A1.
XX
PD   26-FEB-2004.
XX
PF   19-JUL-2003; 2003WO-DE002436.
XX
PR   19-JUL-2002; 2002DE-01033082.
XX
PA   (AMAX-) AMAXA GMBH.
XX
PI   Altrogge L, Males T;
XX
DR   WPI; 2004-238793/22.
XX
PT   New fluorescent proteins, useful as pharmaceuticals, biological markers
PT   and for optical information storage, comprise short-lived but restorable
PT   fluorescence.
XX
PS   Example 1; Fig 1; 55pp; German.
XX
CC   The present invention provides protein and coding sequences of novel
CC   fluorescent proteins. These are useful in pharmaceutical compositions and
CC   as fluorescent markers for gene expression and protein localization in
CC   cellular, developmental or molecular biology, particularly for measuring
CC   time-dependent cellular process such as protein diffusion and transport,
CC   also for optical information storage. The present sequence is a Discosoma
CC   fluorescent protein.
XX
SQ   Sequence 226 AA;

Query Match             96.8%;  Score 1175.5; DB 8; Length 226;
Best Local Similarity   97.8%;  Pred. No. 2.2e-123;
Matches  221;  Conservative    4;  Mismatches    0;  Indels    1;  Gaps    1;

Qy           1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLIKVTKGGPLPFPAWDI 60
                |||| |||||||||||||||||||||||||||||||||||||  || |||||||||||||||
Db           1 MR6SKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHMTVLKVTKGGPLPFPAWDI 60

Qy          61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGCFIY 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LSPQFQYGSKYVKHPADIDYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGCFIY 120

Qy         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKST 180

Qy         181 YMAKK-PVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
                ||||| ||||||||||||||||||||||||||||||||||||||||
Db         181 YMAKKAPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226
```

```
XX
PF   10-APR-2002; 2002US-00121258.
XX
PR   26-FEB-2001; 2001US-00794308.
PR   24-MAY-2001; 2001US-00866538.
XX
PA   (TSIE/) TSIEN R Y.
PA   (CAMP/) CAMPBELL R E.
XX
PI   Tsien RY, Campbell RE;
XX
DR   WPI; 2003-743764/70.
DR   N-PSDB; ADC24128.
XX
PS   Example 1; SEQ ID NO 4; 67pp; English.
XX
CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   containing a vector which comprises (I) operatively linked to an
CC   expression control sequence, and an unit to assay the variant fluorescent
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescence protein produced by (VII), where variant fluorescent protein
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcriptions, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescence sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This is the amino acid sequence of
CC   Discosoma red fluorescent protein variant fast T1.
XX
SQ   Sequence 225 AA;

Query Match             95.6%;  Score 1160; DB 7; Length 225;
Best Local Similarity   96.0%;  Pred. No. 1.2e-121;
Matches  216;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy           1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLIKVTKGGPLPFPAWDI 60
                |||::|||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db           1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFPAWDI 60

Qy          61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGSFIY 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGCFIY 120

Qy         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy         181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
                ||||||||||||||||||||||||||||||||||:|||:||||||
Db         181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225
```

```
RESULT 84
ADC24129

ID   ADC24129 standard; protein; 225 AA.
XX
AC   ADC24129;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Discosoma red fluorescent protein variant fast T1.
XX
KW   Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW   fluorescent protein variant; transcription induction detection;
KW   fluorescence energy resonance transfer; FRET; protein kinase;
KW   protein phosphatase; ion indicator; mutant; mutein; fast T1.
XX
OS   Synthetic.
OS   Discosoma.
XX
PN   US2003059835-A1.
XX
PD   27-MAR-2003.
```

```
Query Match             95.6%;
Best Local Similarity   96.0%;  Conservative
Matches  216;

Qy           1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLIKVTKGGPLPFPAWDI 60
                |||::||||||||||||||||||||||||||||||||||||||
Db           1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEG...

Qy          61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGSFIY 120
Db          61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDDGGVVTVTQDSSLQDGCFIY 120

Qy         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Db         121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy         181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
Db         181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225
```

RESULT 85
ID ADL46206 standard; protein; 225 AA.
XX
AC ADL46206;
XX
DT 20-MAY-2004 (first entry)
XX
DE Discosoma red fluorescent protein; DsRed; fluorescence; red wavelength;
DE oligomerization; tetramerization; immunoassay; hybridization assay.
XX
KW red fluorescent protein; DsRed; fluorescence; red wavelength;
KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS Discosoma sp.
XX
PN WO2003086446-A1.
XX
PD 23-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010879.
XX
PR 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-00209208.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Campbell RE, Baird GS;
XX
DR WPI; 2003-845265/78.
DR N-PSDB; ADL46207.
XX
PT New monomeric and dimeric Anthozoan fluorescent protein variants with
PT reduced propensity to oligomerize, and encoding polynucleotides, useful
PT in molecular biology, e.g. in immunoassays or in tracking protein
PT movement in cells.
XX
PS Disclosure; SEQ ID NO 4; 166pp; English.
XX
CC The invention relates to a polynucleotide sequence encoding a Discosoma
CC red fluorescent protein (DsRed) variant having a reduced propensity to
CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC sequence, where the substitutions result in reduced propensity of the
CC DsRed variant to form tetramers and where the variant displays detectable
CC fluorescence of at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the DsRed
CC variant T1 protein.
XX
SQ Sequence 225 AA;

Query Match   95.6%; Score 1160; DB 7; Length 225;
Best Local Similarity   96.0%;
Matches 216; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            || ||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225

───

RESULT 86
ID ADL46226 standard; protein; 225 AA.
XX
AC ADL46226;
XX
DT 20-MAY-2004 (first entry)
XX
DE Discosoma red fluorescent protein (DsRed) T1 protein with I125R mutation.
XX
KW red fluorescent protein; DsRed; fluorescence; red wavelength;
KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS Discosoma sp.
XX
PN WO2003086446-A1.
XX
PD 23-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010879.
XX
PR 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-00209208.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Campbell RE, Baird GS;
XX
DR WPI; 2003-845265/78.
XX
PT New monomeric and dimeric Anthozoan fluorescent protein variants with
PT reduced propensity to oligomerize, and encoding polynucleotides, useful
PT in molecular biology, e.g. in immunoassays or in tracking protein
PT movement in cells.
XX
PS Disclosure; SEQ ID NO 24; 166pp; English.
XX
CC The invention relates to a polynucleotide sequence encoding a Discosoma
CC red fluorescent protein (DsRed) variant having a reduced propensity to
CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC sequence, where the substitutions result in reduced propensity of the
CC DsRed variant to form tetramers and where the variant displays detectable
CC fluorescence of at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the DsRed
CC variant T1 protein with a I125R mutation.
XX
SQ Sequence 225 AA;

Query Match   95.6%; Score 1153; DB 7; Length 225;
Best Local Similarity   95.0%;
Matches 215; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            || ||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
           |||||||||||||||||||||||||||||||||||||||||
Db   181

Left column:

Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225

XX

AC

ID   ADM13537 standard; protein; 225 AA.

XX

DT   15-JUL-2004 (first entry)

XX

DE   Superfolding coral DsRed protein #1.

XX

KW   GFP; superfolder mutant; mutein; green fluorescent protein;
KW   directed-evolution; folding interference domain; reporter protein;
KW   mutant; DsRed; coral.

XX

OS   Discosoma sp.
OS   Synthetic.

XX

PN   US2004078148-A1.

XX

PD   22-APR-2004.

XX

PF   24-APR-2003; 2003US-00423688.

XX

PR   24-APR-2002; 2002US-00132067.

XX

PA   (WALD/) WALDO G S.

XX

PI   Waldo GS;

XX

DR   WPI; 2004-340059/31.
DR   N-PSDB; ADM13536.

XX

PT   New modified green fluorescent protein, useful as a reporter in
PT   expression studies.

XX

PS   Claim 18; SEQ ID NO 41; 46pp; English.

XX

CC

CC   The invention relates to a green fluorescent protein (GFP) that comprises
CC   at least 80% identity to the GFP superfolding mutant appearing as
CC   ADM13535 that has at least one amino acid substitution selected for
CC   e.g., a substitution at position 30 that is an arginine or a conservative
CC   variant of arginine and measurable fluorescence activity. Also included
CC   are directed-evolution for generating an enhanced folding variant of a
CC   target polypeptide (comprising mutating a polynucleotide encoding
CC   polypeptide of interest to generate a library of mutated polynucleotides,
CC   linking mutated polynucleotides to polynucleotide encoding folding
CC   interference domain to form fusion protein (FP) constructs, expressing
CC   FP, and selecting FP that display optimal folding activity in relation to
CC   FP comprising wild-type polypeptide and folding interference domain, thus
CC   identifying polypeptide with enhanced folding activity), an enhanced
CC   folding variant of a fluorescent protein generated by the method, an
CC   enhanced folding variant of a chromophoric protein generated by the
CC   method, enhancing folding of a polypeptide comprising multiple domains
CC   (involving joining a first domain of the polypeptide to a poorly folding
CC   domain, to form a fusion protein, mutating the first domain, detecting an
CC   increase in the amount of activity generated by a first mutated fusion
CC   protein in comparison to a fusion protein comprising a wild-type first
CC   domain and the poorly folding polypeptide domain, thus identifying a
CC   first domain with enhanced folding, joining a second domain of the
CC   polypeptide to the first mutated fusion protein to form a second fusion
CC   protein, mutating the second domain, and detecting an increase in the
CC   amount of activity generated by a second mutated fusion protein in
CC   comparison to a fusion protein comprising the wild-type second domain and
CC   the first mutated fusion protein, thus identifying a target polypeptide
CC   with multiple domains that have enhanced folding), an isolated nucleic
CC   acid encoding the GFP, an expression vector comprising the nucleic acid
CC   and a host cell comprising the vector. The DsRed protein from coral is
CC   also considered for mutation using the above method. The enhanced folding
CC   variants are useful as reporter proteins to express the report level of a
CC   protein. The method is useful for directed-evolution of generating

Right column:

CC   enhanced folding variant of target polypeptide and for improving folding
CC   and solubility of a target protein. The present sequence represents a
CC   coral DsRed enhanced folding variant. NOTE: The authors refer to SEQ ID
CC   38-42 as SEQ ID 1-5 throughout the specification and never refer to SEQ ID
CC   ID 1-37 by number, it is obvious that SEQ ID 38-42 are meant to be SEQ ID
CC   1-5.

XX

SQ   Sequence 225 AA;

Query Match   93.2%;   Score 1132;   DB 8;   Length 225;
Best Local Similarity   92.4%;   Pred. No. 1.8e-118;
Matches 208; Conservative   7; Mismatches   10; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEPMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
      | |::|:||||||||| ||||||||||||||||||:|||||||||||||||||||
Db   1 MESSEDVIKEPMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQNVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
      |||||||||||||||||||||| ||||||||:||||||| |||||:||||||:||| |:||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNFEDGGVATVTQDSSLEDGCLIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
      ||||||||||| |||||:||||||||||:|||||:||||||:|||||||||||||:|||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEPSTERLYPRDGVLKGDIHKALKLKDGGHYLVDIKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
      |||||||||||||||||||:||:|||||||||||| |||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITNHNEDYTIVEQYERAEGRHHLFL 225

RESULT 88

ID   ADC24131 standard; protein; 226 AA.

XX

AC   ADC24131;

XX

DT   18-DEC-2003 (first entry)

XX

DE   Discosoma red fluorescent protein variant dimer2.

XX

KW   Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW   fluorescent protein variant; transcription induction detection;
KW   fluorescence energy resonance transfer; FRET; Protein kinase;
KW   protein phosphatase; ion indicator; mutant; dimer2; mutein.

XX

OS   Synthetic.
OS   Discosoma.

XX

PN   US2003059835-A1.

XX

PD   27-MAR-2003.

XX

PF   10-APR-2002; 2002US-00121258.

XX

PR   26-FEB-2001; 2001US-00794308.
PR   24-MAY-2001; 2001US-00866538.

XX

PA   (TSIE/) TSIEN R Y.
PA   (CAMP/) CAMPBELL R E.

XX

PI   Tsien RY, Campbell RE;

XX

DR   WPI; 2003-743764/70.
DR   N-PSDB; ADC24130.

XX

PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.

XX

PS   Claim 12; SEQ ID NO 6; 67pp; English.

XX

CC

CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced

CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramer. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   containing a vector which comprises (I) operatively linked to an
CC   expression control sequence, and an unit to assay the variant fluorescent
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescent protein produced by (VII), where variant fluorescent protein
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcription, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This is the amino acid sequence of
CC   Discosoma red fluorescent protein variant dimer2.
XX
SQ   Sequence 226 AA;

Query Match   92.3%; Score 1121; DB 7; Length 226;
Best Local Similarity   92.8%;
Matches 207; Conservative   7; Pred. No. 3.1e-117; Mismatches   9; Indels   0; Gaps   0;

Qy   3 SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS   62
Db   4 SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS   63

Qy   63 PQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKV 122
Db   64 PQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV 123

Qy 123 KFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYM 182
Db 124 KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYM 183

Qy 183 AKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
Db 184 AKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFL 226

RESULT 89
ID   ADL46208
AC   ADL46208;
DT   20-MAY-2004 (first entry)
DE   Discosoma red fluorescent protein (DsRed) variant dimer2 protein.
KW   red fluorescence; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
OS   Discosoma sp.
PN   WO2003086446-A1.
PD   23-OCT-2003.
PF   09-APR-2003; 2003WO-US010879.
PR   10-APR-2002; 2002US-00121258.

ADL46208
ID   ADL46208 standard; protein; 226 AA.

PR   29-JUL-2002; 2002US-00209208.
PA   (REGC) UNIV CALIFORNIA.
PI   Tsien RY, Campbell RE, Baird GS;
DR   WPI; 2003-845265/78.
DR   N-PSDB; ADL46209.
XX
    Disclosure; SEQ ID NO 6; 166pp; English.

    The invention relates to a polynucleotide sequence encoding a Discosoma
    red fluorescent protein (DsRed) variant having a reduced propensity to
    oligomerize. The protein variant comprises one or more amino acid
    substitutions at the AB and/or AC interface(s) of the wild-type DsRed
    sequence, where the substitutions result in reduced propensity of the
    DsRed variant to form tetramers and where the variant displays detectable
    fluorescence of at least one red wavelength. The composition and methods
    are useful in producing red fluorescent proteins having reduced
    propensity for oligomerization, especially tetramerization. The protein
    may be used in molecular biology and in other scientific applications,
    such as in immunoassays or hybridization assays, or in tracking the
    movement of proteins in cells. This sequence corresponds to the DsRed
    variant dimer2 protein.
XX
    New monomeric and dimeric Anthozoan fluorescent protein variants with
    reduced propensity to oligomerize, and encoding polynucleotides, useful
    in molecular biology, e.g. in immunoassays or in tracking protein
    movement in cells.
XX
SQ   Sequence 226 AA;

Query Match   92.3%; Score 1121; DB 7; Length 226;
Best Local Similarity   92.8%;
Matches 207; Conservative   7; Pred. No. 3.1e-117; Mismatches   9; Indels   0; Gaps   0;

Qy   3 SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS   62
Db   4 SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS   63

Qy   63 PQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKV 122
Db   64 PQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV 123

Qy 123 KFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYM 182
Db 124 KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYM 183

Qy 183 AKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
Db 184 AKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFL 226

RESULT 90
AAY99838
ID   AAY99838 standard; protein; 225 AA.
AC   AAY99838;
DT   19-SEP-2000 (first entry)
DE   Humanised anthozoa novel fluorescent protein hybrid drFP583/dmFP592.
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling; hybrid.
OS   Discosoma sp; "red".
OS   Synthetic.
PN   WO200034326-A1.
PD   15-JUN-2000.
XX

```
PF   10-DEC-1999; 99WO-US029473.
XX
PR   11-DEC-1998; 98US-00210330.
PR   14-OCT-1999; 99US-0018529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
DR   WPI; 2000-423381/36.
DR   N-PSDB; AAA48748.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 20; Page 78; 86pp; English.
XX
CC   The present sequence is humanised drFP583 and dmFP592. Amplified wild-type
CC   coding region fragments from drFP583 and dmFP592 were digested with
CC   restriction endonucleases and combined together to produce a hybrid
CC   construct. drFP583 is a full-length cDNA encoding a novel fluorescent
CC   protein (nFP) from Discosoma sp. "red", a non-bioluminescent species of
CC   the Class Anthozoa. The hybrid sequence was changed to optimise the
CC   codons for expression of the protein in mammalian cells. Fluorescent
CC   protein can be used in fluorescent labeling, a useful tool for marking a
CC   protein, cell or organism of interest. Unlike other markers used in
CC   protein labeling, such as beta-galactosidase and luciferase, fluorescent
CC   proteins do not require an exogenous cofactor or substrate. Methods
CC   involving fluorescent proteins are also less laborious and less difficult
CC   to control than the traditional methods of fluorescent labeling, where a
CC   protein of interest is purified and then covalently conjugated to a
CC   fluorophore derivative. Novel fluorescent proteins isolated from species
CC   of the Class Anthozoa can be used as markers for gene expression and
CC   protein localization studies, and in fluorescence resonance energy
CC   transfer (FRET) reactions. They may have improved properties and better
CC   suitability for larger excitations compared to prior art fluorescent
CC   proteins such as green fluorescent protein
XX
SQ   Sequence 225 AA;

Query Match             92.2%;  Score 1119;  DB 3;  Length 225;
Best Local Similarity   91.6%;  Pred. No. 5.1e-117;
Matches  206;  Conservative  12;  Mismatches  7;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||| |:||||||||| :||||:||||||||||| :||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||:|||||||| |||||||||||||||||||||||||||:||||:||||||
Db   61 LSPQFQYGSKYYVKHPADIDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          :|||||||||||||||::|:.|||||:||||||||||||:||:||:||:||||||||||
Db  121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 91

AAE28835
ID   AAE28835 standard; protein; 225 AA.
XX
AC   AAE28835;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583/dmFP592 (6/9Q) protein.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
```

```
KW   fluorescence activated cell sorting application; fluorescence timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; dmFP592 protein.
XX
OS   Discosoma sp.
XX
PN   WO200268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270993P.
PR   04-DEC-2001; 2001US-00005922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   N-PSDB; AAD46280.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page 72-73; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. drFP583/dmFP592 (6/9Q) protein of the invention
XX
SQ   Sequence 225 AA;

Query Match             92.2%;  Score 1119;  DB 5;  Length 225;
Best Local Similarity   91.6%;  Pred. No. 5.1e-117;
Matches  206;  Conservative  12;  Mismatches  7;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||| |:||||||||| :||||:||||||||||| :||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||:|||||||| |||||||||||||||||||||||||||:||||:||||||
Db   61 LSPQFQYGSKYYVKHPADIDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          :|||||||||||||||::|:.|||||:||||||||||||:||:||:||:||||||||||
Db  121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 92

AAY99839
ID   AAY99839 standard; protein; 225 AA.
XX
AC   AAY99839;
XX
DT   19-SEP-2000 (first entry)
XX
DE   Humanised Anthozoa novel fluorescent protein hybrid drFP583/dmFP592-2G.
XX
```

KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling; hybrid; mutant; mutein.
XX
OS   Discosoma sp; "red".
OS   Synthetic.
XX
PN   WO20003426-A1.
XX
PD   15-JUN-2000.
XX
PD   10-DEC-1999.
XX
PF   11-DEC-1998;   98US-00210330.
PR   14-OCT-1999;   99US-00418529.
XX
PA   (CION-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
PS   Claim 6; Page 79; 86pp; English.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
CC   The present sequence is derived from humanised hybrid protein
CC   drFP583/dmFP592. drFP583 is a full-length cDNA encoding a novel
CC   fluorescent protein (nFP) from Discosoma sp. "red", a non-bioluminescent
CC   species of the Class Anthozoa. The hybrid sequence was changed to
CC   optimise the codons for expression of the protein in mammalian cells. The
CC   present sequence contains two substitutions, K15Q and T217S, which were
CC   incorporated by random mutagenesis. Fluorescent proteins can be used in
CC   fluorescent labeling, a useful tool for marking a protein, cell or
CC   organism of interest. Unlike other markers used in protein labeling, such
CC   as beta-galactosidase and luciferase, fluorescent proteins do not require
CC   an exogenous cofactor or substrate. Methods involving fluorescent
CC   proteins are also less laborious and less difficult to control than the
CC   traditional methods of fluorescent labeling, where a protein of interest
CC   is purified and then covalently conjugated to a fluorophore derivative.
CC   Novel fluorescent proteins isolated from species of the Class Anthozoa
CC   can be used as markers for gene expression and protein localization
CC   studies, and in fluorescence resonance energy transfer (FRET) reactions.
CC   They may have improved properties and better suitability for larger
CC   excitations compared to prior art fluorescent proteins such as green
CC   fluorescent protein
XX
SQ   Sequence 225 AA;

Query Match   91.5%; Score 1111; DB 3; Length 225;
Best Local Similarity   90.7%;
Matches 204; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||:|||||||||||||||||:||||||||:|||||||||||||||
Db   1 MSCSKNVIKEFMRFQVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNFEDGGVVTVSQDSSLKDGCFIY 120

Qy   121 KVKPIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            :|||:||:||||||:||:||:|:|:||||||||:|:||||||||||:||||||||||||||
Db   121 EVKPIGVNFPSDGPVMQRRTRGWEASSRLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERSEGRHHLFL 225

ID   AAE28928
AAE28928 standard; protein; 225 AA.
XX
AC   AAE28928;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583/dmFP592 (6/90) mutant protein (K15Q+T217S).
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filler;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; dmFP592 protein;
KW   mutant; mutein.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 15
FT   /note= "Wild-type Lys substituted with Gln"
FT   Misc-difference 217
FT   /note= "wild-type Thr substituted with Ser"
XX
PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PD   20-FEB-2002; 2002WO-US005749.
XX
PF   21-FEB-2001; 2001US-0270909832.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   WPI; 2002-691654/74.
XX
CC   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
CC   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
CC   analyte detection assays or fluorescence activated cell sorting
CC   applications.
XX
PT   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescent timers. The present
CC   sequence is Discosoma sp. drFP583/dmFP592 (6/90) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583/dmFP592 (6/90) wild-type protein shown
CC   as SEQ ID NO:12 (AAE28835) in page 72-73 of the specification
XX
SQ   Sequence 225 AA;

Query Match   91.5%; Score 1111; DB 5; Length 225;
Best Local Similarity   90.7%;
Matches 204; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||:|||||||||||||||||:||||||||:|||||||||||||||
Db   1 MSCSKNVIKEFMRFQVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNFEDGGVVTVSQDSSLKDGCFIY 120

```
Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||::||:||||:|||||||||||:||:|||:|::||||||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         ||||||||||||||||||||||||||||||||||||:|||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERSEGRHHLFL 225
```

RESULT 94
AAE28929

ID   AAE28929 standard; protein; 223 AA.
XX
AC   AAE28929;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583/dmFP592 (6/9Q) mutant protein (Sdel+Cdel+K5E+K9T).
XX
OS   Discosoma sp.
OS   Synthetic.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; Second messenger detector; drFP583 protein; dmFP592 protein;
KW   mutant; mutein.
XX
FH   Key   Location/Qualifiers
FH
FT   Misc-difference 5
FT   /note= "Wild-type Lys substituted with Glu"
FT   Misc-difference 9
FT   /note= "Wild-type Lys substituted with Thr"
XX
PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-0006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
XX
PT   New nucleic acid encoding a non-aggregating Cnidarian chromo- or
PT   fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in fluorescence activated cell sorting applications,
CC   in protease cleavage assays or as fluorescence timers. The present
CC   sequence is Discosoma sp. drFP583/dmFP592 (6/9Q) mutant protein of the
CC   invention. Note: This sequence is not shown in the specification, but is
CC   derived from Discosoma sp. drFP583/dmFP592 (6/9Q) wild-type protein shown
CC   as SEQ ID NO:12 (AAE28835) in page 72-73 of the specification.
.
SQ   Sequence 223 AA;

Query Match   91.1%; Score 1106; DB 5; Length 223;
Best Local Similarity   91.4%; Pred. No. 1.5e-115;
Matches 203; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

Qy   4 SKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSP 63
         :|||||||||||||||||||||||||||||||||:|||||:|||||:||||||||:||||
Db   2 SENVITEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDILSP 61

Qy   64 QFQYGSKVYVKHPADIIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVK 123
         ||||||||||||||||:|||||||||||||||||:|||||||||||:||||:||||||:|
Db   2 QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVVNFEDGGVVTVSQDSSLKDGCFIYEVK 121

Qy   124 FIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMA 183
         |||||||||||||||::||:||||:|||||||||||:|||:|||:|::||||||||||||
Db   122 FIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSIYMA 181

Qy   184 KKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         ||||||||||||||||||||||||||||||||:|||||||
Db   182 KKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERSEGRHHLFL 223

RESULT 95
AAY99840

ID   AAY99840 standard; protein; 225 AA.
XX
AC   AAY99840;
XX
DT   19-SEP-2000 (first entry)
XX
DE   Humanised Anthozoa novel fluorescent protein hybrid drFP583/dmFP592-Q3.
XX
OS   Synthetic.
OS   Discosoma sp. "red".
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent protein labeling; hybrid; mutant; mutein.
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029473.
XX
PR   11-DEC-1998; 98US-00210330.
PR   14-OCT-1999; 99US-00418529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
DR   WPI; 2000-423381/36.
DR   N-PSDB; AAA48750.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 20; Page 81-82; 86pp; English.
XX
CC   The present sequence is derived from humanised hybrid protein
CC   drFP583/dmFP592. drFP583 is a full-length cDNA encoding a novel
CC   fluorescent protein (nFP) from Discosoma sp. "red", a non-bioluminescent
CC   species of the Class Anthozoa. The hybrid sequence was changed to
CC   optimise the codons for expression of the protein in mammalian cells. The
CC   present sequence contains three substitutions, K15Q, K83M and T217S,
CC   which were incorporated by random mutagenesis. Fluorescent proteins can
CC   be used in fluorescent labeling, a useful tool for marking a protein,
CC   cell or organism of interest. Unlike other markers used in protein
CC   labeling, such as beta-galactosidase and luciferase, fluorescent proteins
CC   do not require an exogenous cofactor or substrate. Methods involving
CC   fluorescent proteins are also less laborious and less difficult to
CC   control than the traditional methods of fluorescent labeling, where a
CC   protein of interest is purified and then covalently conjugated to a
```

CC fluorophore derivative. Novel fluorescent proteins isolated from species
CC of the Class Anthozoa can be used as markers for gene expression and
CC protein localization studies, and in fluorescence resonance energy
CC transfer (FRET) reactions. They may have improved properties and better
CC suitability for larger excitations compared to prior art fluorescent
CC proteins such as green fluorescent protein
XX
SQ Sequence 225 AA;

Query Match   91.0%; Score 1105; DB 3; Length 225;
Best Local Similarity 90.2%; Pred. No. 1.9e-115;
Matches 203; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||:|||||||||:|||||||||||||||||||||||||:||||||||||||||||||||
Db    1 MSCSKNVIKEFMRFQVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYMKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||::|:|||||||||||||||||||:||:|:||||||||||||||||
Db  121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||:||||||||||:||||||||||:|||||:||||||||||
Db  181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERSEGRHHLFL 225
```

RESULT 96
AC AAE28927
ID AAE28927 standard; protein; 225 AA.
XX
AC AAE28927;
XX
DT 27-DEC-2002 (first entry)
XX
DE Discosoma sp. drFP583/dmFP592 (6/90) mutant protein (K15Q+K83M+T217S).
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescence timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; drFP583 protein; dmFP592 protein;
KW mutant; mutein.
XX
OS Discosoma sp.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Misc-difference   15
FT   /note= "Wild-type Lys substituted with Gln"
FT Misc-difference   83
FT   /note= "Wild-type Lys substituted with Met"
FT Misc-difference   217
FT   /note= "Wild-type Thr substituted with Ser"
XX
PN WO20020268459-A2.
XX
PD 06-SEP-2002.
XX
PF 20-FEB-2002; 2002WO-US005749.
XX
PR 21-FEB-2001; 2001US-0270983P.
PR 04-DEC-2001; 2001US-0006922.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov S, Lukyanov K, Yanushevich Y, Savistsky A, Fradkov A;
XX
DR WPI; 2002-691654/74.

PT fluorophore derivative. Novel fluorescent mutant
PT of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT analyte detection assays or fluorescence activated cell sorting
PT applications.
XX
PS Disclosure; Page; 80pp; English.
XX
CC The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays and their mutants, as markers in
CC recombinant DNA applications, as sunscreens or filters, in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC invention. Note: This sequence is not shown in the specification, but is
CC derived from Discosoma sp. drFP583/dmFP592 (6/90) wild-type protein shown
CC as SEQ ID NO:12 (AAE28835) in page 72-73 of the specification
XX
SQ Sequence 225 AA;

Query Match   91.0%; Score 1105; DB 5; Length 225;
Best Local Similarity 90.2%; Pred. No. 1.9e-115;
Matches 203; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||:|||||||||:|||||||||||||||||||||||||:||||||||||||||||||||
Db    1 MSCSKNVIKEFMRFQVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYMKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||::|:|||||||||||||||||||:||:|:||||||||||||||||
Db  121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||:||||||||||:||||||||||:|||||:||||||||||
Db  181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERSEGRHHLFL 225
```

RESULT 97
AC AAY92894
ID AAY92894 standard; protein; 230 AA.
XX
AC AAY92894;
XX
DT 12-SEP-2000 (first entry)
XX
DE Discosoma sp. "magenta" fluorescent protein dmFP592.
XX
KW dmFP592; fluorescent protein; non-bioluminescent; label; marker;
KW fluorescence resonance energy transfer reaction; FRET.
XX
OS Discosoma sp.
XX
PN WO200034324-A1.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US029412.
XX
PR 11-DEC-1998; 98US-00210330.
PR 19-NOV-1999; 99US-00444341.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
XX
DR WPI; 2000-423379/36.
DR N-PSDB; AAA28958.

RESULT 98
ID   AAB01625 standard; protein; 230 AA.
XX
AC   AAB01625;
XX
DT   12-DEC-2000 (first entry)
XX
DE   Discosoma sp. magenta fluorescence protein dmFP592.
XX
KW   Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;
KW   zFP506; zFP538; drFP583; dsFP483; asFP600; dgFP512; dmFP592.
XX
OS   Discosoma sp.
XX
CC   This is a novel fluorescent protein (dmFP592) from Discosoma sp.
CC   "magenta", a non-bioluminescent species of the genera Anthozoa. The
CC   dmFP592. It was isolated using the Aequoria victoria green fluorescent
CC   fluorescent colour is observed under usual white light or ultraviolet
CC   light. It has an absorbance maximum at 573 nm and an emmision maximum at
CC   593 nm. The polynucleotides of the invention can be used to produce the
CC   proteins recombinantly, and as a source of primers and probes for
CC   identifying related proteins. The fluorescent proteins have applications
CC   in fluorescent labeling, as fluorescent markers for gene expression and
CC   protein localization studies, and in fluorescence resonance energy
CC   transfer (FRET) reactions. They may have improved properties and better
CC   suitability for larger excitations, compared to prior art fluorescent
CC   proteins such as green fluorescent protein from Aequorea victoria, which
CC   has an excitation maximum at 395 nm, a second excitation peak at 475 nm
CC   and an emission maximum at 510 nm
XX
SQ   Sequence 230 AA;

Query Match   89.4%; Score 1085.5; DB 3; Length 230;
Best Local Similarity   88.1%; Pred. No. 3.1e-113;
Matches 199; Conservative 17; Mismatches   9; Indels   1; Gaps   1;

Qy   1 MRSSKNVIKEPMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MSCSKNVIKEPMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||:||:|||||||||||||||||||||||||||:|||||||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKP-VQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||::||||||||||||||||||||||||||||:||||||||:|
Db   181 YMVKKPSVQLPGYYYVVDSKLDMTSHNEDYTVVEQYEKTQGRHHPFI 226

RESULT 99
ID   AAE00378 standard; protein; 230 AA.
XX
AC   AAE00378;
XX
DT   11-SEP-2003 (revised)
DT   19-JUN-2001 (first entry)
XX
DE   Discosoma sp. magenta chromo/fluorescent protein, dmFP592 (NFP-9).
XX
KW   Anthozoa; Chromoprotein; fluorescent protein; dmFP592; NFP-9; sunscreen;
KW   analyte detection assay; selectable marker; recombinant DNA application;
KW   biosensor; pH indicator; invivo marker; selective filter.
XX
OS   Discosoma sp; magenta.
XX
CC   The present sequence is Discosoma sp. magenta fluorescent protein
CC   dmFP592. It was isolated using the Aequoria victoria green fluorescent
CC   protein (GFP) sequence, which was used to design PCR primers which might
CC   isolate other fluorescent proteins from a number of species of Anthozoa.
CC   These were Anemonia sulcata, Clavularia sp., Zoanthus sp., Discosoma sp.
CC   and Anemonia majano. The cDNA obtained was then screened in the search
CC   for sequences encoding fluorescent proteins. The other proteins found in
CC   this manner were cFP484, zFP506, zFP538, drFP583, dsFP483, asFP600,
CC   dgFP512 and amFP486. These proteins can be used as fluorescent labels
CC   (for gene expression and protein localisation studies and in fluorescence
CC   resonance energy transfer (FRET) studies) in place of fluorophore
CC   derivatives and luciferases, as these involve laborious processes and the
CC   latter require cofactors. They can also be used in place of GFP, which is
CC   too stable to be useful when studying short-term or repetitive events
XX
SQ   Sequence 230 AA;

Query Match   89.4%; Score 1085.5; DB 3; Length 230;
Best Local Similarity   88.1%; Pred. No. 3.1e-113;
Matches 199; Conservative 17; Mismatches   9; Indels   1; Gaps   1;

Qy   1 MRSSKNVIKEPMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MSCSKNVIKEPMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||:||:|||||||||||||||||||||||||||:|||||||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKP-VQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||::||||||||||||||||||||||||||||:||||||||:|
Db   181 YMVKKPSVQLPGYYYVVDSKLDMTSHNEDYTVVEQYEKTQGRHHPFI 226

PS   Claim 20; Page 55-56; 60pp; English.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma species,
PT   useful for fluorescent labeling and as markers.
XX
PN   WO20003452-6-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029405.
XX
PR   11-DEC-1998; 98US-0021030.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
XX
DR   WPI; 2000-42451/36.

PS   Claim 3; Page 71-72; 73pp; English.
XX
PT   ,Novel method for identifying a DNA sequence encoding fluorescent proteins
PT   from non-bioluminescent Anthozoa which are useful for fluorescent
PT   labeling and as markers.
XX
PN   WO20012715-0-A2.
XX
PD   19-APR-2001.
XX
PF   13-OCT-2000; 2000WO-US028477.
XX
PR   14-OCT-1999; 99US-00418529.
PR   15-OCT-1999; 99US-00418917.
PR   15-OCT-1999; 99US-00418922.
PR   19-NOV-1999; 99US-00444338.
PR   19-NOV-1999; 99US-00444341.
PR   09-DEC-1999; 99US-00457556.

PR 09-DEC-1999; 99US-00457898.
PR 09-DEC-1999; 99US-00458144.
PR 09-DEC-1999; 99US-00458477.
PR 10-DEC-1999; 99WO-US029405.
PR 14-JUN-2000; 2000US-0211607P.
PR 14-JUN-2000; 2000US-0211609P.
PR 14-JUN-2000; 2000US-0211626P.
PR 14-JUN-2000; 2000US-0211627P.
PR 14-JUN-2000; 2000US-0211687P.
PR 14-JUN-2000; 2000US-0211766P.
PR 14-JUN-2000; 2000US-0211880P.
PR 14-JUN-2000; 2000US-0211880P.
PR 14-JUN-2000; 2000US-0212070P.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Terskikh A;
XX
DR WPI; 2001-266409/27.
DR N-PSDB; AAD03618.
XX
PT An Anthozoa chromo- or fluorescent protein (P1) present in an environment
PT other than its natural environment, useful as a label in analyte
PT detection assays and as a selectable marker in recombinant DNA
PT applications.
XX
PS Claim 8; Fig 9; 69pp; English.
XX
CC The present sequence is a Discosoma sp. magenta chromo/fluorescent
CC protein, dmFPS92 (NFP-9). NFP-9 is present in an environment other than
CC its natural environment and has an absorbance maximum ranging from 375nm
CC to 775nm and more usually from 560nm to 590nm and emission maximum
CC ranging from 395nm and more usually from 580 to 610nm. The
CC chromoproteins or fluorescent proteins are useful as labels in analyte
CC detection assays, as selectable markers in recombinant DNA applications,
CC as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
CC and as in vivo markers in animals. They are also useful in sunscreens and
CC as selective filters. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 230 AA;

Query Match   89.4%; Score 1085.5; DB 4; Length 230;
Best Local Similarity   88.1%; Pred. No. 3.1e-113;
Matches 199; Conservative 17; Mismatches   9; Indels   1; Gaps   1;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
           ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||:||
Db   1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI   60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY   120
           |||||||||||||||||||||||||:||||||||||||||||||||||||:|||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY   120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180
           ||||||||||||||||||||||||||||||||||||||||:|:|||||||||||||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI   180

Qy   181 YMAKKP-VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL   225
           |||||| |||||||||||:|||||||||||||||||||:|||||:
Db   181 YMVKKPSVQLPGYYYVDSKLDMYSHNEDYTVVEQYEKTQGRHPFI   226

KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO20027070-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2001US-02732227P.
XX
PR 02-MAR-2001; 2001US-02732227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU-) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PS Example 19; Page 498; 510pp; English.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 230 AA;

Query Match   89.4%; Score 1085.5; DB 5; Length 230;
Best Local Similarity   88.1%; Pred. No. 3.1e-113;
Matches 199; Conservative 17; Mismatches   9; Indels   1; Gaps   1;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
           ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||:||
Db   1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI   60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNNFEDGGVVTVTQDSSLQDGCFIY   120
           |||||||||||||||||||||||||:||||||||||||||||||||||||:|||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIY   120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI   180
           ||||||||||||||||:|:||||||||||||||||||||:|:|||||||||||||||||:
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI   180

AC ABP70038;
XX
ID ABP70038 standard; protein; 230 AA.
XX
RESULT 100
ABP70038
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour facilitating molecule (CFM) related sequence #SEQ ID 240.
XX

```
Qy    181 YMAKKP-VQLPGYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||| ||||||||||||||||||||||||||||||||||| |:||
Db    181 YMVKKPSVQLPGYYVDSKLDMTSHNEDYTVVEQYEKTQGRHHPFI 226

RESULT 101
ID    ABP96651
AC    ABP96651 standard; protein; 230 AA.
XX
DT    29-MAY-2003  (first entry)
XX
DE    Red fluorescent protein (DsRed) protein SEQ ID NO:40.
XX
KW    Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
KW    att site; integrase; recombinase; Aces; gene therapy; transgenic animal;
KW    platform artificial chromosome expression system.
XX
OS    Discosoma sp.
OS    Synthetic.
XX
PN    WO20029705-A2.
XX
PD    05-DEC-2002.
XX
PF    30-MAY-2002; 2002WO-US017452.
PR    30-MAY-2001; 2001US-0294758P.
PR    21-MAR-2002; 2002US-0368091P.
XX
DR    WPI; 2003-140461/13.
DR    N-PSDB; ACC44649.
XX
PA    (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX
PI    Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
PI    Stewart S, Shellard J;
XX
PT    Novel eukaryotic chromosome comprising one or many att sites which
PT    permits site-directed integration in the presence of lambda-integrase,
PT    useful for site-specific recombination-directed integration of DNA of
PT    interest.
XX
PS    Example 1; Page 192-193; 272pp; English.
XX
CC    The present invention describes a eukaryotic chromosome (I) comprising
CC    one or several att sites, where an att site is heterologous to the
CC    chromosome, and permits site-directed integration in the presence of
CC    lambda-integrase. Also described: (1) a platform artificial chromosome
CC    expression system (Aces) (II) comprising several sites that participate
CC    in recombinase catalysed recombination; and (2) a method (M1) for
CC    introducing a heterologous nucleic acid into a platform artificial
CC    chromosome. (I) can be used in gene therapy. (M1) is useful for
CC    introducing a heterologous nucleic acid molecule into a platform
CC    artificial chromosome, preferably an Aces. (II) is useful for producing a
CC    transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
CC    mammal) by introducing (II) by cell fusion, lipid-mediated transfection,
CC    microprojectile bombardment or direct DNA transfer into an embryonic
CC    cell, preferably a stem cell or an embryo. (II) comprises a heterologous
CC    nucleic acid that encodes a therapeutic product which is useful for
CC    making a library of Aces comprising random portions of a genome. ACC44612
CC    to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
CC    exemplification of the present invention
XX
SQ    Sequence 230 AA;

Query Match    89.4%; Score 1085.5; DB 6; Length 230;
Best Local Similarity  88.1%; Pred. No. 3.1e-113;
Matches  199; Conservative  17; Mismatches  9; Indels  1; Gaps  1;

Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
```

```
Db      1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy     61 LSPQFQYGSKYVKHPADIDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             :|:|||||:|||||||||||||:|::|||||||||:|||||||||||||||||:||||
Db     61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
             :|:|||||||::||||:|||||:||||||||||||||||:|||:||||||||||||||||
Db    121 EVKFIGVNFPSDGPVWQRRRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy    181 YMAKKP-VQLPGYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
             ||||| ||||||||||||||||||||||||||||||||||| |:||
Db    181 YMVKKPSVQLPGYYVDSKLDMTSHNEDYTVVEQYEKTQGRHHPFI 226

RESULT 102
ID    AAE00375
AC    AAE00375 standard; protein; 205 AA.
XX
DT    11-SEP-2003  (revised)
DT    19-JUN-2001  (first entry)
XX
DE    Discosoma sp. red chromo/fluorescent protein, drFP583 (NFP-6).
XX
KW    Anthozoa; Chromoprotein; fluorescent protein; drFP583; NFP-6; sunscreen;
KW    analyte detection assay; selectable marker; recombinant DNA application;
KW    biosensor; pH indicator; invivo marker; selective filter.
XX
OS    Discosoma sp; red.
XX
FH    Key            Location/Qualifiers
FT    Misc-difference 150..151
FT                    /note= "Encoded by TTGTATCCTCGTGATGGCGTGTTGAAAAGGAGAGATT
FT                    CATAAGGCTCTGAAGCTGAAAGACGGTGGT"
XX
PN    WO20012715O-A2.
XX
PD    19-APR-2001.
XX
PF    13-OCT-2000; 2000WO-US028477.
XX
PR    14-OCT-1999; 99US-00418529.
PR    15-OCT-1999; 99US-00418917.
PR    15-OCT-1999; 99US-00418922.
PR    19-NOV-1999; 99US-00444338.
PR    19-NOV-1999; 99US-00444341.
PR    09-DEC-1999; 99US-00457556.
PR    09-DEC-1999; 99US-00457755.
PR    09-DEC-1999; 99US-00457898.
PR    09-DEC-1999; 99US-00458144.
PR    09-DEC-1999; 99US-00458477.
PR    10-DEC-1999; 99US-00459405.
PR    14-JUN-2000; 2000WO-US029405.
PR    14-JUN-2000; 2000US-02116070P.
PR    14-JUN-2000; 2000US-02116609P.
PR    14-JUN-2000; 2000US-02116626P.
PR    14-JUN-2000; 2000US-02116627P.
PR    14-JUN-2000; 2000US-02116870P.
PR    14-JUN-2000; 2000US-02117660P.
PR    14-JUN-2000; 2000US-02118800P.
PR    14-JUN-2000; 2000US-02118880P.
PR    14-JUN-2000; 2000US-02120770P.
XX
PA    (CLON-) CLONTECH LAB INC.
XX
PI    Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Terskikh A;
XX
PT    An Anthozoa chromo- or fluorescent protein (P1) present in an environment
PT    other than its natural environment, useful as a label in analyte
```

PT   detection assays and as a selectable marker in recombinant DNA
PT   applications.
XX
PS   Claim 8; Fig 6; 69pp; English.
XX
SQ   Sequence 205 AA;

Query Match   87.1%; Score 1058; DB 4; Length 205;
Best Local Similarity   88.4%; Pred. No. 3.3e-110;
Matches 199; Conservative 2; Mismatches 4; Indels 20; Gaps 1;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        :.||||||||||||||||||||||||||||||                  ||||||||||
Db  121 KSSSLALNFPSDGPVMQKKTMGWEASTERL------------------GHYLVEFKSI 160

Qy  181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  161 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 205
```

RESULT 103
ID   AAE00384 standard; protein; 205 AA.
XX
AC   AAE00384;
XX
DT   19-JUN-2001 (first entry)
XX
DE   Discosoma sp. red drFP583 (NFP-6) mutant E5up (V105A).
XX
KW   Anthozoa; Chromoprotein; fluorescent protein; drFP583; NFP-6; sunscreen;
KW   analyte detection assay; selectable marker; recombinant DNA application;
KW   biosensor; pH indicator; invivo marker; selective filter; mutant; mutein.
XX
OS   Discosoma sp; red.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference   105
FT   /note= "Wild type Val substituted with Ala"
XX
PN   WO020127150-A2.
XX
PD   19-APR-2001.
XX
PF   13-OCT-2000; 2000WO-US028477.
XX
PR   14-OCT-1999; 99US-00418529.
PR   15-OCT-1999; 99US-00418917.
PR   15-OCT-1999; 99US-00418922.
PR   19-NOV-1999; 99US-00444338.
PR   19-NOV-1999; 99US-00444341.
PR   09-DEC-1999; 99US-00457556.
PR   09-DEC-1999; 99US-00457898.
PR   09-DEC-1999; 99US-00458144.
PR   09-DEC-1999; 99US-00458477.
PR   10-DEC-1999; 99WO-US029405.
PR   14-JUN-2000; 2000US-02116097.
PR   14-JUN-2000; 2000US-02116609P.
PR   14-JUN-2000; 2000US-02116626P.
PR   14-JUN-2000; 2000US-02116627P.
PR   14-JUN-2000; 2000US-02116687P.
PR   14-JUN-2000; 2000US-02117766P.
PR   14-JUN-2000; 2000US-02118800P.
PR   14-JUN-2000; 2000US-02118880P.
PR   14-JUN-2000; 2000US-02120700P.
XX
DR   WPI; 2001-266409/27.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Terskikh A;
XX
PA   (CLON-) CLONTECH LAB INC.
XX

CC   The present sequence is a Discosoma sp. red chromo/fluorescent protein,
CC   drFP583 (NFP-6) mutant E5up (V105A). NFP-6 is present in an environment
CC   other than its natural environment and has an absorbance maximum ranging
CC   from 250nm to 750nm and more usually from 540nm to 580nm and emission
CC   maximum ranging from 275nm to 775nm and more usually from 565 to 605nm.
CC   The chromoproteins or fluorescent proteins are useful as labels in
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications, as biosensors in prokaryotic and eukaryotic cells e.g. as
CC   pH indicator and as invivo markers in animals. They are also useful in
CC   sunscreens and as selective filters. Note: The present sequence is not
CC   shown in the specification, but derived from drFP583 (NFP-6) sequence
CC   (AAE00375) shown in figure 6
XX

PT   other than its natural environment, useful as a label in analyte
PT   detection assays and as a selectable marker in recombinant DNA
PT   applications.
XX
PS   Disclosure; Page; 69pp; English.
XX
SQ   Sequence 205 AA;

Query Match   86.8%; Score 1054; DB 4; Length 205;
Best Local Similarity   88.0%; Pred. No. 9.3e-110;
Matches 198; Conservative 2; Mismatches 5; Indels 20; Gaps 1;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        :.||||||||||||||||||||||||||||                    ||||||||||
Db  121 KSSSLALNFPSDGPVMQKKTMGWEASTERL------------------GHYLVEFKSI 160

Qy  181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  161 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 205
```

RESULT 104
ID   AAE00383 standard; protein; 205 AA.
XX
AC   AAE00383;

CC   An Anthozoa chromo- or fluorescent protein (P1) present in an environment

XX
DT   19-JUN-2001 (first entry)
XX
DE   Discosoma sp. red drFP583 (NFP-6) mutant E8 (N42H).
XX
KW   Anthozoa; Chromoprotein; fluorescent protein; drFP583; NFP-6; sunscreen;
KW   analyte detection assay; selectable marker; recombinant DNA application;
KW   biosensor; pH indicator; invivo marker; selective filter; mutant; mutein.
XX
OS   Discosoma sp; red.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 42
FT   /note= "wild type Asn substituted with His"
XX
PN   WO200127150-A2.
XX
PD   19-APR-2001.
XX
PF   13-OCT-2000; 2000WO-US028477.
XX
PR   14-OCT-1999;   99US-00418529.
PR   15-OCT-1999;   99US-00418917.
PR   15-OCT-1999;   99US-00418922.
PR   19-NOV-1999;   99US-00444338.
PR   19-NOV-1999;   99US-00444341.
PR   09-DEC-1999;   99US-00457556.
PR   09-DEC-1999;   99US-00457898.
PR   09-DEC-1999;   99US-00458144.
PR   09-DEC-1999;   99US-00458477.
PR   10-DEC-1999;   99WO-US029405.
PR   14-JUN-2000;   2000US-02160P.
PR   14-JUN-2000;   2000US-0211607P.
PR   14-JUN-2000;   2000US-0211609P.
PR   14-JUN-2000;   2000US-0211626P.
PR   14-JUN-2000;   2000US-0211627P.
PR   14-JUN-2000;   2000US-0211687P.
PR   14-JUN-2000;   2000US-0211766P.
PR   14-JUN-2000;   2000US-0211880P.
PR   14-JUN-2000;   2000US-0211888P.
PR   14-JUN-2000;   2000US-0212070P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Terskikh A;
XX
DR   WPI; 2001-266409/27.
XX

PT   An Anthozoa chromo- or fluorescent protein, useful as a label in analyte
PT   detection assays and as a selectable marker in recombinant DNA
PT   applications.
XX
PS   Disclosure; Page; 69pp; English.
XX

CC   The present sequence is a Discosoma sp. red chromo/fluorescent protein,
CC   drFP583 (NFP-6) mutant E8 (N42H). NFP-6 is present in an environment
CC   other than its natural environment and has an absorbance maximum ranging
CC   from 250nm to 750nm more usually from 540nm and emission maximum ranging
CC   from 275nm to 775nm and more usually from 565 to 605nm.
CC   The chromoproteins or fluorescent proteins are useful as labels in
CC   analyte detection assays, as selectable markers in recombinant DNA
CC   applications, as biosensors in prokaryotic and eukaryotic cells e.g. as
CC   pH indicator and as in vivo markers in animals. They are also useful in
CC   sunscreens and as selective filters. Note: The present sequence is not
CC   shown in the specification, but derived from drFP583 (NFP-6) sequence
CC   (AAB00375) shown in figure 6
CC
SQ   Sequence 205 AA;
.

Query Match   86.7%; Score 1053; DB 4; Length 205;
Best Local Similarity 88.0%; Pred. No. 1.2e-109;
, Matches 198; Conservative 3; Mismatches 4; Indels 20; Gaps 1;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||| :|||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHHTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||||||||||||| :|||||||||||||||||||||||||
Db   121 KSSSLALNFPSDGPVMQKKTMGWEASTERL------------------GHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   161 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 205

RESULT 105
ADC24133
ID   ADC24133 standard; protein; 225 AA.
XX
AC   ADC24133;
XX
DT   18-DEC-2003 (first entry)
DE   Discosoma red fluorescent protein variant mRFP1.
XX
KW   Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW   fluorescent protein variant; transcription induction detection;
KW   fluorescence energy resonance transfer; FRET; protein kinase;
KW   protein phosphatase; ion indicator; mutant; mRFP1; mutein.
XX
OS   Synthetic.
OS   Discosoma.
XX
PN   US2003059835-A1.
XX
PD   27-MAR-2003.
XX
PF   10-APR-2002; 2002US-00121258.
XX
PR   26-FEB-2001; 2001US-00794308.
PR   24-MAY-2001; 2001US-0086538.
XX
PA   (TSIE/) TSIEN R Y.
PA   (CAMP/) CAMPBELL R E.
XX
PI   Tsien RY, Campbell RE;
XX
DR   WPI; 2003-743764/70.
DR   N-PSDB; ADC24132.
XX

PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.
XX
PS   Disclosure; SEQ ID NO 8; 67pp; English.
XX

CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerize, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (SI) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   containing a vector which comprises (II) operatively linked to an
CC   expression control sequence, and an unit to assay fluorescence of the
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescent protein produced by (VII), where variant fluorescent
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide

CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcriptions, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This is the amino acid sequence of
CC   Discosoma red fluorescent protein variant mRFP1.
XX
SQ   Sequence 225 AA;

Query Match   84.1%; Score 1021; DB 7; Length 225;
Best Local Similarity   86.9%; Pred. No. 5.6e-106;
Matches 192; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           |:|::||||||||||||||||| ||||||||||||||||| | |||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||:||||||||||:||||||||||||||||||||||||||||||||||||:|||
Db   61 LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           |||::|:.|||||||||||||||||||||||:|||:|||:||||:||:|||||||:|:||:|
Db   121 KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
           |||||||||||||||||||||||||||||||||||||:|||||
Db   181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 221

RESULT 106
ID   ADL46210 standard; protein; 225 AA.
XX
AC   ADL46210;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Discosoma red fluorescent protein (DsRed) variant mRFP1 protein.
XX
KW   red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
PN   WO2003086446-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-0209208.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Telen RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
DR   N-PSDB; ADL46211.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein

PT   movement in cells.
XX
PS   Disclosure; SEQ ID NO 8; 166pp; English.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific application,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   variant mRFP1 protein.
XX
SQ   Sequence 225 AA;

Query Match   84.1%; Score 1021; DB 7; Length 225;
Best Local Similarity   86.9%; Pred. No. 5.6e-106;
Matches 192; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           |:|::||||||||||||||||| ||||||||||||||||| | |||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
           ||||||||||:||||||||||:||||||||||||||||||||||||||||||||||||:|||
Db   61 LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
           |||::|:.|||||||||||||||||||||||:|||:|||:||||:||:|||||||:|:||:|
Db   121 KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
           |||||||||||||||||||||||||||||||||||||:|||||
Db   181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 221

RESULT 107
ID   ADQ59566 standard; protein; 225 AA.
XX
AC   ADQ59566;
XX
DT   07-OCT-2004 (first entry)
XX
DE   Discosoma monomerised red fluorescent protein mRFP1.
XX
KW   optically detectable signal; protein fragment complementation assay; PCA;
KW   molecular interaction detection; biomolecular interaction detection;
KW   fluorescent protein; multi-colour PCA; drug discovery; target validation;
KW   high-throughput screening; high-content screening; pathway mapping;
KW   drug mechanism-of-action study; biosensor; diagnostic;
KW   red fluorescent protein; spectral shift; mRFP1.
XX
OS   Discosoma.
XX
PN   US2004137528-A1.
XX
PD   15-JUL-2004.
XX
PF   01-DEC-2003; 2003US-0724178.
XX
PR   02-FEB-1998; 98US-00017412.
PR   07-FEB-2000; 2000US-0999464.
PR   24-MAY-2002; 2002US-00154758.
PR   29-JAN-2003; 2003US-00353090.
PR   09-APR-2003; 2003US-0461133P.
XX
PA   (MICH/) WATSON MICHNICK S W.

PA   (MACD/)) MACDONALD M L.
PA   (LAME/) LAMERDIN J.
XX
PI   Watson Michnick SW, Macdonald ML, Lamerdin J;
XX
DR   WPI; 2004-533363/51.
DR   N-PSDB; ADQ59565.
XX
PT   Composition useful in protein fragment complementation assays for drug
PT   discovery and high-throughput screening, comprising complementary
PT   fragments of protein or mutant protein, generating optically detectable
PT   signal when associated.
XX
XX
PS   Disclosure; SEQ ID NO 16; 34pp; English.
XX
CC   The invention describes a composition (I) comprising complementary
CC   fragments of a protein or mutant protein, where the fragments generate an
CC   optically detectable signal when associated, and each of the mutant
CC   protein fragments is fused to a separate molecule. Also described are:
CC   protein fragment complementation assays (PCA) for detection of molecular
CC   interactions, involving reassembling separate fragments from an optically
CC   detectable protein, and detecting the reassembly by units of optically
CC   detectable protein, and detecting the activity of the optically detectable
CC   reconstitution of activity of the optically detectable protein, where the
CC   reassembly of the fragments is operated by the interaction of molecular
CC   domains fused to each fragment, and is independent of other molecular
CC   processes; detecting biomolecular interaction, involving selecting an
CC   appropriate optically detectable protein, effecting fragmentation of the
CC   optically detectable protein such that the fragmentation results in
CC   reversible loss of protein function, fusing or attaching fragments of the
CC   optically detectable protein separately to other molecules, reassociating
CC   the protein fragments through interactions of the molecules that are
CC   fused or attached to the fragments, and detecting the resulting optical
CC   signal; designing and engineering of PCAs based on fluorescent protein;
CC   and a method and composition for the construction of multi-color PCAs.
CC   (I) is useful in PCAs and other assays for drug discovery, target
CC   validation, high-throughput screening, high-content screening, target
CC   mapping, drug mechanism-of-action studies, biosensors and diagnostics.
CC   (I) is useful for engineering different colour PCAs for a variety of
CC   applications in biology and biotechnology. This is the amino acid
CC   sequence of a Discosoma monomerised red fluorescent protein from which
CC   PCA fragments can be isolated.
XX
SQ   Sequence 225 AA;

Query Match   84.1%; Score 1021; DB 8; Length 225;
Best Local Similarity   86.9%; Pred. No. 5.6e-106;
Matches 192; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         | :: |||||||||||||||||| |||||||||||||||||| |||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI   60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||.||||||||||| ||||||||||||||||||||||||||||||||| |||
Db   61 LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||| | ||||||||||||||||||||||.||.||.||||| .||||||||| . :.
Db 121 KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRH 221
         |||||||||||. . ||.|||||||||||||||||||||||
Db 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 221

RESULT 108
AEA54910
ID   AEA54910 standard; protein; 747 AA.
XX
AC   AEA54910;
XX
DT   25-AUG-2005 (first entry)
XX

DE   Fluorescence-related EGFP-caspase-3 DEVD linker-mRFP(x2) fusion protein.
XX
KW   fluorescence; protein localization; cellular transport; organelle;
KW   mitochondria; fusion protein; enhanced green fluorescent protein; EGFP;
KW   caspase-3.
XX
OS   Unidentified.
OS   Chimeric.
XX
XX
PN   WO2005054464-A1.
XX
PD   16-JUN-2005.
XX
PF   03-DEC-2003; 2004WO-JP018437.
XX
PR   03-DEC-2003; 2003JP-00404472.
PR   27-JAN-2004; 2004JP-00018344.
XX
DR   WPI; 2005-444979/45.
DR   N-PSDB; AEA54911.
XX
PA   (RIKE) RIKEN KK.
PA   (MEDI-) MEDICAL & BIOLOGICAL LAB CO LTD.
XX
PI   Miyawaki A, Kogure T, Hama H, Kinjo M, Saito K, Karasawa S;
PI   Araki T;
XX
PT   Novel fluorescent protein existing as monomer and isolated from Fungia
PT   and Montipora species, useful for labeling and analyzing location and
PT   movement of intracellular organelles.
XX
PS   Example 10; SEQ ID NO 49; 218pp; Japanese.
XX
CC   The invention relates to a novel fluorescent protein existing as a
CC   monomer and comprising any one of SEQ ID No. 1-29 (odd SEQ ID numbers)
CC   given in the specification, or any one of the sequences chosen from SEQ
CC   ID No. 1-29 (odd SEQ ID numbers), with one or more amino acid
CC   substitutions, deletions and/or additions, and having fluorescent
CC   characteristics. The fusion protein of the invention may be useful for
CC   analyzing the location or movement of a protein in a cell in vivo and for
CC   labeling and analyzing the location or movement of intracellular
CC   organelles, such as mitochondria. The current sequence is that of the
CC   fluorescence-related enhanced green fluorescent protein (EGFP)-caspase-3
CC   DEVD linker-mRFP(x2) fusion protein of the invention.
XX
SQ   Sequence 747 AA;

Query Match   84.1%; Score 1021; DB 9; Length 747;
Best Local Similarity   86.9%; Pred. No. 3.3e-105;
Matches 192; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         | :: |||||||||||||||||| |||||||||||||||||| |||||||||||||||
Db 269 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 328

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||.||||||||||| ||||||||||||||||||||||||||||||||| |||
Db 329 LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 388

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||| | ||||||||||||||||||||||.||.||.||||| .||||||||| . :.
Db 389 KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 448

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRH 221
         |||||||||||. . ||.|||||||||||||||||||||||
Db 449 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 489

RESULT 109
ADL46281
ID   ADL46281 standard; protein; 225 AA.
XX
AC   ADL46281;

XX
DT 20-MAY-2004 (first entry)
XX
DE Discosoma red fluorescent protein (DsRed) variant mRFP1.1 coding sequence.
XX
KW red fluorescent protein; DsRed; fluorescence; red wavelength;
KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS Discosoma sp.
OS Synthetic.
XX
PN WO200308446-A1.
XX
PD 23-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010879.
XX
PR 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-00209208.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Campbell RE, Baird GS;
XX
DR WPI; 2003-845265/78.
XX
PT New monomeric and dimeric Anthozoan fluorescent protein variants with
PT reduced propensity to oligomerize, and encoding polynucleotides, useful
PT in molecular biology, e.g. in immunoassays or in tracking protein
PT movement in cells.
XX
PS Claim 70; SEQ ID NO 79; 166pp; English.
XX
CC The invention relates to a polynucleotide sequence encoding a Discosoma
CC red fluorescent protein (DsRed) variant having a reduced propensity to
CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC sequence, where the substitutions result in reduced propensity of the
CC DsRed variant to form tetramers and where the variant displays detectable
CC fluorescence of at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the DsRed
CC variant mRFP1.1 protein which contains a Q66M and a T147S mutation.
XX
SQ Sequence 225 AA;

Query Match   83.4%; Score 1012; DB 7; Length 225;
Best Local Similarity   86.0%; Pred. No. 5.8e-105;
Matches 190; Conservative 6; Mismatches 25; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            | ||||:|||||||||||||||||:|||||||||||||||||| ||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVVVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVQDSSLQDGCFIY 120
            |||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||
Db   61 LSPQFMYGSKAVVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVQDSSLQDGEFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||| |||||||||||||||||||||||||:||||||||||||||||||||||||| |:
Db   121 KVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTY 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
            |||||||||:|| | ||||||||||||||||||||||| |||
Db   181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 221

Search completed: January 11, 2006, 02:08:28
Job time : 81 secs

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

```
                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2006   Compugen Ltd.
```

Run on: January 11, 2006, 02:05:28 ; Search time 16 Seconds
                                                (without alignments)
                                                1353.048 Million cell updates/sec

Title: US-10-006-922A-12
Perfect Score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..........EDYTIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
                    Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 69%
                 Maximum Match 100%
                 Listing first 500 summaries

Database :   PIR_80:*
                    1:   pir1:*
                    2:   pir2:*
                    3:   pir3:*
                    4:   pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result   Query
No.   Score   Match   Length DB   ID   Description
------ -----   -----   ------ --   --   -----------

No matches found

Search completed: January 11, 2006, 02:10:16
Job time : 16 secs

OM protein - protein search, using sw model

Run on: January 11, 2006, 02:04:58 ; Search time .73 Seconds

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Title:
Perfect Score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..........EDYTIVEQYERTEGRHHLFL 225
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues

Database : UniProt_05.80:*
1: uniprot_sprot.*
2: uniprot_trembl:*

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 69%
Maximum Match 100%
Listing first 500 summaries

(without alignments)
2174.574 Million cell updates/sec

Total number of hits satisfying chosen parameters: 7

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	225	1	RFP_DISSP	
2	1196	98.5	236	2	Q66ND7_9CNID	discosoma s
3	1193	98.3	225	2	Q6KF85_9CNID	discosoma s
4	1191	98.1	236	2	Q66ND8_9CNID	discosoma s
5	1085.5	89.4	230	2	Q9GTJ7_9CNID	discosoma s
6	985	81.1	226	2	Q53G8_9CNID	discosoma s
7	972	80.1	226	2	Q53G7_9CNID	discosoma s

ALIGNMENTS

RESULT 1
RFP_DISSP
ID RFP_DISSP STANDARD; PRT; 225 AA.
AC Q9U6Y8;
DT 01-FEB-2005 (Rel. 46, Created)
DT 01-FEB-2005 (Rel. 46, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Red fluorescent protein drFP583 (DsRed).
OS Discosoma sp. (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=86600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99436614; PubMed=10504696; DOI=10.1038/13657;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).

RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX PubMed=11101896; DOI=10.1038/81992;
RA Wall M.A., Socolich M., Ranganathan R.;
RT "The structural basis for red fluorescence in the tetrameric GFP
RT homolog DsRed.";
RL Nat. Struct. Biol. 7:1133-1138(2000).

RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND SUBUNIT.
RX PubMed=11209050; DOI=10.1073/pnas.98.2.462;
RA Yarbrough D., Wachter R.M., Kallio K., Matz M.V., Remington S.J.;
RT "Refined crystal structure of DsRed, a red fluorescent protein from
RT coral, at 2.0-A resolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:462-467(2001).
-!- FUNCTION: Thought to play a role in photoprotection of the coral's
CC resident symbiont microalgae's photosystems from photoinhibition
CC caused by high light levels found near the surface of coral reefs.
CC In deeper water, the fluorescence may be to convert blue light
CC into longer wavelengths more suitable for use in photosynthesis by
CC the microalgal symbionts.
-!- BIOPHYSICOCHEMICAL PROPERTIES:
CC Absorption:
CC Abs(max)=558 nm;
CC Note=Exhibits a smaller absorbance peak at 494 nm. The broad
CC fluorescence emission spectrum peaks at 583 nm;
-!- SUBUNIT: Homotetramer.
-!- PTM: Contains a chromophore consisting of modified amino acid
CC residues. The chromophore is formed by autocatalytic backbone
CC condensation between Xaa-N and Gly-(N+2), oxidation of Tyr-(N+1)
CC to didehydrotyrosine, and formation of a double bond to the alpha-
CC amino nitrogen of residue Xaa-N. Maturation of the chromophore
CC requires nothing other than molecular oxygen.
-!- BIOTECHNOLOGY: Fluorescent proteins have become a useful and
CC ubiquitous tool for making chimeric proteins, where they function
CC as a fluorescent protein tag. Typically they tolerate N- and C-
CC terminal fusion to a broad variety of proteins. They have been
CC expressed in most known cell types and are used as a noninvasive
CC fluorescent marker in living cells and organisms. They enable a
CC wide range of applications where they have functioned as a cell
CC lineage tracer, reporter of gene expression, or as a measure of
CC protein-protein interactions.
-!- SIMILARITY: Belongs to the GFP family.
-!- This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
DR EMBL; AF168419; AAF03369.1; -; mRNA.
DR PDB; 1G7K; X-ray; A/B/C/D=2-225.
DR PDB; 1GGX; X-ray; A/B/C/D=1-225.
DR InterPro; IPR011584; GFP_related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Chromophore; Luminescence; Photoprotein.
KW 3D-structure; Chromophore; Luminescence; Photoprotein.
FT MOD_RES 66 68 (2)-2,3-didehydrotyrosine.
FT CROSSLNK 67 67 2-iminomethyl-5-imidazolinone (Gln-Gly).
SQ SEQUENCE 225 AA; 25931 MW; FBF9A536978F689 CRC64;

Query Match 100.0%; Score 1214; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e-96; Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy 61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

```
Qy  121 KVKFIGVNFPSDGPVWQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVWQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 2
```
ID    Q66ND7_9CNID PRELIMINARY;     PRT;     236 AA.
AC    Q66ND7;
DT    25-OCT-2004 (TrEMBLrel. 28, Created)
DT    25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT    25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE    Enhanced red fluorescent protein R+.
OS    Discosoma sp. RC-2004.
OC    Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC    Discosomatidae; Discosoma.
OX    NCBI_TaxID=289055;
RN    [1]
RP    NUCLEOTIDE SEQUENCE.
RA    Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT    "Cloning of Anthozoan Fluorescent Protein Genes.";
RL    Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AY679107; AAU04444.1; -; mRNA.
DR    GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR    InterPro; IPR011584; GFP related.
DR    InterPro; IPR000786; Green_fl_protein.
DR    Pfam; PF01353; GFP; 1.
DR    PRINTS; PR01229; GFLUORESCENT.
DR    ProDom; PD013756; Green_fl_protein; 1.
SQ    SEQUENCE   236 AA;   27032 MW;   BB3844BE06829EF0 CRC64;

Query Match             98.5%;  Score 1196;  DB 2;  Length 236;
Best Local Similarity   98.2%;  Pred. No. 3.9e-95;
Matches   221;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVWQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVWQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 3
```
ID    Q6KF85_9CNID PRELIMINARY;     PRT;     225 AA.
AC    Q6KF85;
DT    05-JUL-2004 (TrEMBLrel. 27, Created)
DT    05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT    05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE    Orange fluorescent protein FP586.
OS    Discosoma sp. JW-2002.
OC    Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC    Discosomatidae; Discosoma.
OX    NCBI_TaxID=208461;
RN    [1]
RP    NUCLEOTIDE SEQUENCE.
RA    Wiedenmann J., Girod A.;
RL    Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF545828; AAQ11987.1; -; mRNA.
DR    HSSP; P42212; 1B9C.
DR    GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR    InterPro; IPR011584; GFP related.
DR    InterPro; IPR000786; Green_fl_protein.
DR    Pfam; PF01353; GFP; 1.
DR    PRINTS; PR01229; GFLUORESCENT.
DR    ProDom; PD013756; Green_fl_protein; 1.
SQ    SEQUENCE   225 AA;   25791 MW;   E151D0E497AA23FA CRC64;
```

RESULT 4
```
ID    Q66ND8_9CNID PRELIMINARY;     PRT;     236 AA.
AC    Q66ND8;
DT    25-OCT-2004 (TrEMBLrel. 28, Created)
DT    25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT    25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE    Red fluorescent protein R1.
OS    Discosoma sp. RC-2004.
OC    Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC    Discosomatidae; Discosoma.
OX    NCBI_TaxID=289055;
RN    [1]
RP    NUCLEOTIDE SEQUENCE.
RA    Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT    "Cloning of Anthozoan Fluorescent Protein Genes.";
RL    Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AY679106; AAU04443.1; -; mRNA.
DR    GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR    InterPro; IPR011584; GFP related.
DR    InterPro; IPR000786; Green_fl_protein.
DR    Pfam; PF01353; GFP; 1.
DR    PRINTS; PR01229; GFLUORESCENT.
DR    ProDom; PD013756; Green_fl_protein; 1.
SQ    SEQUENCE   236 AA;   27042 MW;   BB3844BCD6834EF3 CRC64;

Query Match             98.1%;  Score 1191;  DB 2;  Length 236;
Best Local Similarity   97.8%;  Pred. No. 1.1e-94;
Matches   220;  Conservative   2;  Mismatches   3;  Indels   0;  Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVWQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db  121 KVKFIGVNFPSDGPVWQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180
```

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225

RESULT 5
Q9GTU7_9CNID

ID   Q9GTU7_9CNID PRELIMINARY;   PRT;   230 AA.
AC   Q9GTU7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 25, Last annotation update)
DE   Red fluorescent protein.
GN   Name=FP593;
OS   Discosoma sp. SSAL-2000.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=137428;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=20434599; PubMed=10981720; DOI=10.1016/S0014-5793(00)01895-0;
RA   Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA   Lukyanov S.A.;
RT   "Novel fluorescent protein from Discosoma coral and its mutants
RT   possesses a unique far-red fluorescence.";
RL   FEBS Lett. 479:127-130(2000).
DR   EMBL; AF272711; AAG16224.1; -; mRNA.
HSSP; Q9U6Y8; 1GGX.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
SQ   SEQUENCE   230 AA; 26370 MW; 5215B1B436D67E51 CRC64;

Query Match   89.4%; Score 1085.5; DB 2; Length 230;
Best Local Similarity   88.1%; Pred. No. 1.3e-85;
Matches 199; Conservative 17; Mismatches   9; Indels   1; Gaps   1;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLKDGCFIY 120

Qy   121 KVKPIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKP-VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||| ||:|||||||||||||||:||:|:|||||:|:|||||:
Db   181 YMVKKPSVQLPGYYYVDSKLDMTSHNEDYTIVEQYEKTQGRHPFI 226

RESULT 6
Q5S3G8_9CNID

ID   Q5S3G8_9CNID PRELIMINARY;   PRT;   226 AA.
AC   Q5S3G8;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Fluorescent protein raspberry.
OS   Discosoma sp. LW-2004.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=301246;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15556995; DOI=10.1073/pnas.0407752101;
RA   Wang L., Jackson W.C., Steinbach P.A., Tsien R.Y.;
RT   "Evolution of new nonantibody proteins via iterative somatic

RT   hypermutation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:16745-16749(2004).
EMBL; AY786536; AAV65486.1; -; mRNA.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   226 AA; 25513 MW; 159A8FBE2D453A5B CRC64;

Query Match   81.1%; Score 985; DB 2; Length 226;
Best Local Similarity   85.3%; Pred. No. 6.3e-77;
Matches 185; Conservative   6; Mismatches 26; Indels   0; Gaps   0;

Qy   5 KNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQ 64
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 EEVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTARLKVTKGGPLPFAWDILSPQ 65

Qy   65 FQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF 124
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 CMYGSKGYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGALKGEMKMRLKLKDGGHYDAEVKTTYMAK 185

Qy   125 IGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 184
            ||||||||||||||||||||:||:||||||||||||||||||:|:||||||||||||
Db   126 RGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEMKMRLKLKDGGHYDAEVKTTYMAK 185

Qy   185 KPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
            ||:|||||||||||||||:||||||||||||||:|||
Db   186 KPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 222

RESULT 7
Q5S3G7_9CNID

ID   Q5S3G7_9CNID PRELIMINARY;   PRT;   226 AA.
AC   Q5S3G7;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Fluorescent protein plum.
OS   Discosoma sp. LW-2004.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=301246;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15556995; DOI=10.1073/pnas.0407752101;
RA   Wang L., Jackson W.C., Steinbach P.A., Tsien R.Y.;
RT   "Evolution of new nonantibody proteins via iterative somatic
RT   hypermutation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:16745-16749(2004).
EMBL; AY786537; AAV65487.1; -; mRNA.
SMR; Q5S3G7; 8-222.
DR   GO; GO:0006091; P:generation of precursor metabolites and energy; IEA.
DR   InterPro; IPR011584; GFP_related.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein.
SQ   SEQUENCE   226 AA; 25590 MW; 6DF38CAD2AB28BED CRC64;

Query Match   80.1%; Score 972; DB 2; Length 226;
Best Local Similarity   83.4%; Pred. No. 8.4e-76;
Matches 181; Conservative   8; Mismatches 28; Indels   0; Gaps   0;

Qy   5 KNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQ 64
            ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 EEVIKEFMRFKEHMEGSVAGHEFEIEGEGEGRPYEGTQTARLKVTKGGPLPFAWDILSPQ 65

Qy   65 FQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF 124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||:|
Db   66 IMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKV 125

```
Qy   125 IGVNFPSDGPVMQKKTMGWEASTERLIYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 184
             |||||||||||||||||||||||||: :|| | :|| :|||| :||||
Db   126 RGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEMKMRLRLKDGGHYDAEVKTTYMAK 185

Qy   185 KPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
             ||||||| |||||||||||||||||||| |||
Db   186 KPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 222
```

Search completed: January 11, 2006, 02:09:49
Job time : 74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2006, 02:06:44 ; Search time 23 Seconds
(without alignments)
808.784 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 69%
Maximum Match 100%
Listing first 500 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	1214	100.0	225	2	US-09-866-538-12	Sequence 12, Appl
2	1214	100.0	225	2	US-09-865-291-12	Sequence 12, Appl
3	1210	99.7	240	2	US-10-152-296-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-866-538-12
; Sequence 12, Application US/0986538
; Patent No. 6852849
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; US-09-866-538-12

Query Match 100.0%; Score 1214; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 3e-132;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 2
US-09-865-291-12
; Sequence 12, Application US/09865291
; Patent No. 6900304
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: ZHANG, Jin
; APPLICANT: TING, Alice
; TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; US-09-865-291-12

Query Match 100.0%; Score 1214; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 3e-132;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 3
US-10-152-296-2
; Sequence 2, Application US/10152296
; Patent No. 6723537
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-0001100US

```
;   CURRENT APPLICATION NUMBER: US/10/152,296
;   CURRENT FILING DATE: 2002-12-10
;   PRIOR APPLICATION NUMBER: US 60/291,871
;   PRIOR FILING DATE: 2001-05-18
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: PatentIn Ver. 2.1
;
;   SEQ ID NO 2
;   LENGTH: 240
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:mammalian
;   OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
;   OTHER INFORMATION: "red" red fluorescent protein (RFP)
US-10-152-296-2

Query Match       99.7%;   Score 1210;   DB 2;   Length 240;
Best Local Similarity   99.6%;   Pred. No. 9.5e-132;
Matches  224;   Conservative   1;   Mismatches   0;   Indels   0;   Gaps   0;

Qy         1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2  VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  61

Qy        61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  121

Qy       121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       122  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  181

Qy       181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
              ||||||||||||||||||||||||||||||||||||||||||||
Db       182  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  226
```

Search completed: January 11, 2006, 02:10:50
Job time : 23 secs

OM protein - protein search, using sw model

Run on: January 11, 2006, 02:07:04 ; Search time 62 Seconds
(without alignments)
1516.316 Million cell updates/sec

Copyright (c) 1993 - 2006 Compugen Ltd.
GenCore version 5.1.6

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 69%
Maximum Match 100%
Listing first 500 summaries

Database :

Published Applications AA_Main:*
1: /cgn2_6/ptcdata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptcdata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptcdata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptcdata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptcdata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptcdata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1214	100.0	225	3	US-09-999-745-67	Sequence 67, Appli
2	1214	100.0	225	3	US-09-866-538-12	Sequence 12, Appl
3	1214	100.0	225	3	US-09-794-308-12	Sequence 12, Appl
4	1214	100.0	225	3	US-09-865-291-12	Sequence 12, Appli
5	1214	100.0	225	4	US-10-006-922-12	Sequence 12, Appl
6	1214	100.0	225	4	US-10-081-864-8	Sequence 8, Appli
7	1214	100.0	225	4	US-10-121-258-1	Sequence 1, Appli
8	1214	100.0	225	5	US-10-315-920-2	Sequence 2, Appli
9	1214	100.0	225	4	US-10-132-067-4	Sequence 4, Appli
10	1214	100.0	225	5	US-10-370-570-56	Sequence 56, Appl
11	1214	100.0	225	5	US-10-406-616-32	Sequence 32, Appl
12	1214	100.0	225	5	US-10-433-640-13	Sequence 13, Appl
13	1214	100.0	225	4	US-10-311-030-7	Sequence 7, Appl
14	1214	100.0	225	3	US-10-045-484-3	Sequence 3, Appl
15	1214	100.0	225	4	US-10-885-988-12	Sequence 12, Appl
16	1214	100.0	225	5	US-10-857-622-22	Sequence 22, Appl
17	1214	100.0	225	5	US-10-656-022-12	Sequence 12, Appl
18	1214	100.0	225	5	US-10-505-486-26	Sequence 26, Appl
19	1214	100.0	225	5	US-10-844-064A-2	Sequence 2, Appli
20	1214	100.0	225	5	US-10-931-304-1	Sequence 1, Appli
21	1214	100.0	487	4	US-10-343-977-1	Sequence 1, Appli
22	1214	100.0	487	5	US-10-343-977-1	Sequence 1, Appli
23	1214	100.0	487	4	US-10-343-977-2	Sequence 2, Appli
24	1214	100.0	506	5	US-10-343-977-2	Sequence 2, Appli
25	1214	100.0	547	4	US-10-343-977-3	Sequence 3, Appli
26	1214	100.0	547	5	US-10-343-977-3	Sequence 3, Appli
27	1210	99.7	226	4	US-10-311-030-9	Sequence 9, Appli
28	1210	99.7	225	4	US-10-152-256-2	Sequence 2, Appli
29	1210	99.7	240	4	US-10-739-656-2	Sequence 2, Appli
30	1210	99.7	545	4	US-10-214-932-52	Sequence 52, Appl
31	1210	99.7	548	4	US-10-214-258-76	Sequence 76, Appl
32	1207	99.4	225	4	US-10-121-258-20	Sequence 20, Appl
33	1207	99.4	225	5	US-10-931-304-20	Sequence 20, Appl
34	1207	99.4	225	4	US-10-315-920-4	Sequence 4, Appli
35	1196	98.5	225	5	US-10-931-304-2	Sequence 2, Appli
36	1196	98.5	236	4	US-10-314-936-4	Sequence 4, Appli
37	1196	98.5	236	6	US-11-021-014-2	Sequence 2, Appli
38	1196	98.5	236	6	US-11-021-014-4	Sequence 4, Appli
39	1195.5	98.5	226	4	US-10-724-178-12	Sequence 12, Appl
40	1191	98.1	225	4	US-10-315-920-6	Sequence 6, Appli
41	1186	97.7	239	4	US-10-442-148A-7	Sequence 7, Appli
42	1186	97.7	239	4	US-10-442-148A-8	Sequence 8, Appli
43	1160	95.6	225	4	US-10-121-258-4	Sequence 4, Appli
44	1160	95.6	225	5	US-10-931-304-4	Sequence 4, Appli
45	1153	95.0	225	5	US-10-931-304-24	Sequence 24, Appl
46	1153	95.0	225	4	US-10-121-258-24	Sequence 24, Appl
47	1132	93.2	225	4	US-10-423-688A-41	Sequence 41, Appl
48	1121	92.3	225	5	US-10-931-304-106	Sequence 106, App
49	1121	92.3	226	4	US-10-121-258-6	Sequence 6, Appli
50	1121	92.3	602	6	US-11-052-001-6	Sequence 6, Appli
51	1119	92.2	225	4	US-10-006-922-44	Sequence 44, Appl
52	1119	92.2	225	4	US-10-081-864-12	Sequence 12, Appl
53	1095	90.2	234	5	US-10-931-304-81	Sequence 81, Appl
54	1095	90.2	476	4	US-10-423-688A-41	Sequence 41, Appl
55	1085.5	89.4	230	4	US-10-006-922-18	Sequence 18, Appl
56	1085.5	89.4	230	5	US-10-161-403-40	Sequence 40, Appl
57	1085.5	89.4	230	6	US-11-006-076-40	Sequence 40, Appli
58	1046	86.2	205	4	US-10-006-922-46	Sequence 46, Appli
59	1021	84.1	225	4	US-10-121-258-8	Sequence 8, Appli
60	1021	84.1	225	4	US-10-724-178-16	Sequence 16, Appli
61	1021	84.1	225	5	US-10-931-304-8	Sequence 8, Appli
62	1014	83.5	225	5	US-10-931-304-108	Sequence 108, App
63	1012	83.4	225	5	US-10-931-304-79	Sequence 79, Appli
64	1007	82.9	225	5	US-10-931-304-104	Sequence 104, App
65	1002	82.5	235	5	US-10-931-304-92	Sequence 92, Appli
66	994	81.9	235	5	US-10-931-304-94	Sequence 94, Appli
67	992	81.7	236	5	US-10-931-304-85	Sequence 85, Appl
68	989	81.5	236	5	US-10-931-304-102	Sequence 102, App
69	988	81.4	236	5	US-10-931-304-83	Sequence 83, Appl
70	987	81.3	236	5	US-10-931-304-100	Sequence 100, App
71	986	81.2	236	5	US-10-931-304-89	Sequence 89, Appl
72	973	80.1	236	5	US-10-931-304-96	Sequence 96, Appl
73	973	80.1	236	5	US-10-931-304-98	Sequence 98, Appl
74	960	79.1	236	5	US-10-931-304-87	Sequence 87, Appl
75	848	69.9	183	4	US-10-724-178-1041	Sequence 1041, Ap

ALIGNMENTS

RESULT 1
US-09-999-745-67

; Sequence 67, Application US/0999745
; Patent No. US2002015712O0A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.

US-09-999-745-67

Query Match
Best Local Similarity 100.0%; Score 1214; DB 3; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pred. No. 6.6e-114;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

RESULT 2
US-09-866-538-12
Sequence 12, Application US/09865538
Publication No. US20030032088A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530-2
CURRENT APPLICATION NUMBER: US/09/866,538
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin version 3.0
SEQ ID NO 12
LENGTH: 225
TYPE: PRT
ORGANISM: Discosoma sp.
US-09-866-538-12

Query Match
Best Local Similarity 100.0%; Score 1214; DB 3; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pred. No. 6.6e-114;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

RESULT 3
US-09-794-308-12
Sequence 12, Application US/09794308
Publication No. US20030170911A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: ZACHARIAS, David
APPLICANT: BAIRD, Geoffrey

TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530
CURRENT APPLICATION NUMBER: US/09/794,308
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin version 3.0
SEQ ID NO 12
LENGTH: 225
TYPE: PRT
ORGANISM: Discosoma sp.
US-09-794-308-12

Query Match
Best Local Similarity 100.0%; Score 1214; DB 3; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pred. No. 6.6e-114;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

RESULT 4
US-09-865-291-12
Sequence 12, Application US/09865291
Publication No. US20030186229A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: TING, Alice
APPLICANT: ZHANG, Jin
TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
FILE REFERENCE: REGEN1550
CURRENT APPLICATION NUMBER: REGEN1550
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patentin version 3.0
SEQ ID NO 12
LENGTH: 225
TYPE: PRT
ORGANISM: Discosoma sp.
US-09-865-291-12

Query Match
Best Local Similarity 100.0%; Score 1214; DB 3; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pred. No. 6.6e-114;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
```

RESULT 5
US-10-006-922-12
; Sequence 12, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1lel Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-12

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 6
US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 7
US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030005983A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(225)
; OTHER INFORMATION: wild-type DsRed
US-10-121-258-1

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
```

```
Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 8
US-10-315-920-2
; Sequence 2, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-315-920-2

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 9
US-10-132-067-4
; Sequence 4, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
; TITLE OF INVENTION: Fluorescence
; FILE REFERENCE: 021362-0006000US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; OTHER INFORMATION: red fluorescent protein (dsRED)
US-10-132-067-4

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 10
US-10-370-570-56
; Sequence 56, Application US/10370570
; Publication No. US20030219717A1
; GENERAL INFORMATION:
; APPLICANT: DAHL, Soren Weis et al.
; FILE REFERENCE: 3759-0130P
; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS
; CURRENT APPLICATION NUMBER: US/10/370,570
; CURRENT FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-370-570-56

Query Match   100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity   100.0%; Pred. No. 6.6e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 11
US-10-406-618-32
; Sequence 32, Application US/10406618
; Publication No. US20030219814A1
; GENERAL INFORMATION:
; APPLICANT: Wan, David Chi-Cheong
; APPLICANT: Ip, Denis Tsz-Ming
; APPLICANT: The Chinese University of Hong Kong

TITLE OF INVENTION: No. US20030219814A1e1 Fluorescent Proteins
FILE REFERENCE: 016285-34-1
CURRENT APPLICATION NUMBER: US/10/406,618
CURRENT FILING DATE: 2003-04-02
PRIOR APPLICATION NUMBER: US 60/387,968
PRIOR FILING DATE: 2002-06-11
PRIOR APPLICATION NUMBER: US 60/370,598
PRIOR FILING DATE: 2002-04-05
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.1

US-10-406-618-32
SEQ ID NO 32
  LENGTH: 225
  TYPE: PRT
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: Description of Artificial Sequence: Spontaneously
  OTHER INFORMATION: fluorescent protein DsRed.

Query Match   100.0%;   Score 1214;   DB 4;   Length 225;
Best Local Similarity 100.0%;   Pred. No. 6.6e-114;
Matches 225;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 12
US-10-433-640-13
Sequence 13, Application US/10433640
Publication No. US20040115792A1
GENERAL INFORMATION:
  APPLICANT: Lichtenberg-Frate, Hella
  TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
  TITLE OF INVENTION: COMPLEX ENVIRONMENTAL CONTAMINATIONS
  FILE REFERENCE: 1487/3
  CURRENT APPLICATION NUMBER: US/10/433,640
  CURRENT FILING DATE: 2003-10-02
  PRIOR APPLICATION NUMBER: PCT/EP01/14610
  PRIOR FILING DATE: 2001-12-12
  PRIOR APPLICATION NUMBER: DE 10061872.3
  PRIOR FILING DATE: 2000-12-12
  NUMBER OF SEQ ID NOS: 24
  SOFTWARE: Patentin version 3.2

US-10-433-640-13
SEQ ID NO 13
  LENGTH: 225
  TYPE: PRT
  ORGANISM: Discosoma sp.

Query Match   100.0%;   Score 1214;   DB 4;   Length 225;
Best Local Similarity 100.0%;   Pred. No. 6.6e-114;
Matches 225;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 13
US-10-311-030-7
Sequence 7, Application US/10311030
Publication No. US20040171107A1
GENERAL INFORMATION:
  APPLICANT: Nelson, David
  APPLICANT: Zamzara, Elize
  APPLICANT: Teien, Roger
  TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
  FILE REFERENCE: 15916-032US1
  CURRENT APPLICATION NUMBER: US/10/311,030
  CURRENT FILING DATE: 2002-12-10
  PRIOR APPLICATION NUMBER: PCT/US01/04625
  PRIOR FILING DATE: 2001-02-13
  PRIOR APPLICATION NUMBER: US 60/184,732
  PRIOR FILING DATE: 2000-02-23
  NUMBER OF SEQ ID NOS: 12
  SOFTWARE: FastSEQ for Windows Version 4.0

US-10-311-030-7
SEQ ID NO 7
  LENGTH: 225
  TYPE: PRT
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: synthetic construct

Query Match   100.0%;   Score 1214;   DB 4;   Length 225;
Best Local Similarity 100.0%;   Pred. No. 6.6e-114;
Matches 225;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 14
US-10-845-484-3
Sequence 3, Application US/10845484
Publication No. US20040248180A1
GENERAL INFORMATION:
  APPLICANT: Bulina, Maria
  APPLICANT: Chudakov, Dmitry
  APPLICANT: Lukyanov, Konstantin
  TITLE OF INVENTION: Mutant Chromophores/Flourophores and
  TITLE OF INVENTION: Methods for Making and Using the Same
  FILE REFERENCE: CLON 092
  CURRENT APPLICATION NUMBER: US/10/845,484
  CURRENT FILING DATE: 2004-05-12
  PRIOR APPLICATION NUMBER: 60/343128
  PRIOR FILING DATE: 2001-12-26

```
; PRIOR APPLICATION NUMBER: PCT/US02/41418
; PRIOR FILING DATE: 2002-12-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
US-10-845-484-3

Query Match             100.0%;  Score 1214;  DB 5;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 6.6e-114;
Matches 225;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 15
US-10-885-988-12
; Sequence 12, Application US/10885988
; Publication No. US20040259165A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/10/885,988
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.0
;
; SEQ ID NO 12
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
US-10-885-988-12

Query Match             100.0%;  Score 1214;  DB 5;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 6.6e-114;
Matches 225;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 16
US-10-656-029-22
; Sequence 22, Application US/10656029
; Publication No. US20050003367A1
; GENERAL INFORMATION:
; APPLICANT: VERTEX PHARMACEUTICALS INC.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RAPID DEVELOPMENT OF
;   SCREENING ASSAYS
; FILE REFERENCE: VPI/02-143WO2
; CURRENT APPLICATION NUMBER: US/10/656,029
; CURRENT FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: 60/408,297
; PRIOR FILING DATE: 2002-09-05
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 22
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
US-10-656-029-22

Query Match             100.0%;  Score 1214;  DB 5;  Length 225;
Best Local Similarity   100.0%;  Pred. No. 6.6e-114;
Matches 225;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 17
US-10-857-622-12
; Sequence 12, Application US/10857622
; Publication No. US20050026234A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: VIOLIN, Jonathan
; APPLICANT: NEWTON, Alexandra
; APPLICANT: TSIEN, Roger
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: Emission Ratiometric Indicators of Phosphorylation By C-Kinase
; FILE REFERENCE: 39754-0891 CP C1CP2
; CURRENT APPLICATION NUMBER: US/10/857,622
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: US 09/865,291
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/396,003
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: US 08/792,553
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: US 594,575
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 12
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
US-10-857-622-12
```

RESULT 18
US-10-505-486-26
; Sequence 26, Application US/10505486
; Publication No. US20050118639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NO 26
; SEQ ID NO 26
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-505-486-26

Query Match　　　　　100.0%;　　Score 1214;　DB 5;　Length 225;
Best Local Similarity　100.0%;　Pred. No. 6.6e-114;
Matches 225;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60

Qy　　61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVKLKVTQDSSLQDGCFIY 120
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVKLKVTQDSSLQDGCFIY 120

Qy　　121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy　　181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
　　　　|||||||||||||||||||||||||||||||||||||||||||||
Db　　181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 19
US-10-844-064A-2
; Sequence 2, Application US200501499944A
; Publication No. US20050149994A1
; GENERAL INFORMATION:
; APPLICANT: Bevis, Brooke
; APPLICANT: Glick, Benjamin
; TITLE OF INVENTION: RAPIDLY MATURING FLUORESCENT PROTEINS AND METHODS FOR USING THE S

; FILE REFERENCE: 092234-9006
; CURRENT APPLICATION NUMBER: US/10/844,064A
; CURRENT FILING DATE: 2004-05-11
; PRIOR APPLICATION NUMBER: PCT/US02/40539
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US 60/341,723
; PRIOR FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma
US-10-844-064A-2

Query Match　　　　　100.0%;　　Score 1214;　DB 5;　Length 225;
Best Local Similarity　100.0%;　Pred. No. 6.6e-114;
Matches 225;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60

Qy　　61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVKLKVTQDSSLQDGCFIY 120
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVKLKVTQDSSLQDGCFIY 120

Qy　　121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy　　181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
　　　　|||||||||||||||||||||||||||||||||||||||||||||
Db　　181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 20
US-10-931-304-1
; Sequence 1, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 1110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(225)
; OTHER INFORMATION: wild-type DsRed
US-10-931-304-1

Query Match　　　　　100.0%;　　Score 1214;　DB 5;　Length 225;
Best Local Similarity　100.0%;　Pred. No. 6.6e-114;
Matches 225;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 21
US-10-343-977-1

; Sequence 1, Application US/10343977
; Publication No. US20040014134A1
; GENERAL INFORMATION:
; APPLICANT: Kuhlemann, Rene
; APPLICANT: Koltermann, Andre
; APPLICANT: Kettling, Ulrich
; APPLICANT: Schwille, Petra
; TITLE OF INVENTION: Dual coloured fluorimetric protease assay
; FILE REFERENCE: 023425us/JH/ml
; CURRENT APPLICATION NUMBER: US/10/343,977
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: PCT/EP01/09112
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Autofluorescent fusion protein derived from GFP variants of
; OTHER INFORMATION: e.g. Aequorea victoria and/or Discona sp.
US-10-343-977-1

```
Query Match             100.0%;  Score 1214;  DB 4;  Length 487;
Best Local Similarity   100.0%;  Pred. No. 1.9e-113;
Matches  225; Conservative    0; Mismatches   0; Indels   0; Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  263 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 322

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  323 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 382

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  383 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 442

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db  443 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 487
```

RESULT 22
US-10-343-977-1

; Sequence 1, Application US/10343977
; Publication No. US20050112682A9
; GENERAL INFORMATION:
; APPLICANT: Kuhlemann, Rene
; APPLICANT: Koltermann, Andre
; APPLICANT: Kettling, Ulrich
; APPLICANT: Schwille, Petra

; TITLE OF INVENTION: Dual coloured fluorimetric protease assay
; FILE REFERENCE: 023425us/JH/ml
; CURRENT APPLICATION NUMBER: US/10/343,977
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: PCT/EP01/09112
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Autofluorescent fusion protein derived from GFP variants of
US-10-343-977-1

```
Query Match             100.0%;  Score 1214;  DB 5;  Length 487;
Best Local Similarity   100.0%;  Pred. No. 1.9e-113;
Matches  225; Conservative    0; Mismatches   0; Indels   0; Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  263 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 322

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  323 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 382

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  383 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 442

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||||||||||||||||||||||||||||||||||||||||||
Db  443 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 487
```

RESULT 23
US-10-343-977-2

; Sequence 2, Application US/10343977
; Publication No. US20040014134A1
; GENERAL INFORMATION:
; APPLICANT: Kuhlemann, Rene
; APPLICANT: Koltermann, Andre
; APPLICANT: Kettling, Ulrich
; APPLICANT: Schwille, Petra
; TITLE OF INVENTION: Dual coloured fluorimetric protease assay
; FILE REFERENCE: 023425us/JH/ml
; CURRENT APPLICATION NUMBER: US/10/343,977
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: PCT/EP01/09112
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Autofluorescent fusion protein derived from GFP variants of
US-10-343-977-2

```
Query Match             100.0%;  Score 1214;  DB 4;  Length 506;
Best Local Similarity   100.0%;  Pred. No. 2e-113;
Matches  225; Conservative    0; Mismatches   0; Indels   0; Gaps   0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  282 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 341

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
```

```
Db    342 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    402 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 461

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
              |||||||||||||||||||||||||||||||||||||||||||||

Db    462 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 506
```

RESULT 24
US-10-343-977-2
; Sequence 2, Application US/10343977
; Publication No. US20050112682A9
; GENERAL INFORMATION:
; APPLICANT: Kuhlemann, Rene
; APPLICANT: Koltermann, Andre
; APPLICANT: Kettling, Ulrich
; APPLICANT: Schwille, Petra
; TITLE OF INVENTION: Dual coloured fluorimetric protease assay
; FILE REFERENCE: 02342Sus6/JH/ml
; CURRENT APPLICATION NUMBER: US/10/343,977
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: PCT/EP01/09112
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 2
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Autofluorescent fusion protein derived from GFP variants of
; OTHER INFORMATION: e.g. Aequorea victoria and/or Discoma sp.
US-10-343-977-2

Query Match   100.0%; Score 1214; DB 5; Length 506;
Best Local Similarity   100.0%; Pred. No. 2e-113;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    282 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 341

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    342 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 401

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    402 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 461

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
             |||||||||||||||||||||||||||||||||||||||||||||

Db    462 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 506
```

RESULT 25
US-10-343-977-3
; Sequence 3, Application US/10343977
; Publication No. US20040014134A1
; GENERAL INFORMATION:
; APPLICANT: Kuhlemann, Rene
; APPLICANT: Koltermann, Andre
; APPLICANT: Kettling, Ulrich
; APPLICANT: Schwille, Petra
; TITLE OF INVENTION: Dual coloured fluorimetric protease assay
; FILE REFERENCE: 02342Sus6/JH/ml
; CURRENT APPLICATION NUMBER: US/10/343,977
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: PCT/EP01/09112

; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 3
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Artificial sequence
US-10-343-977-3

Query Match   100.0%; Score 1214; DB 4; Length 547;
Best Local Similarity   100.0%; Pred. No. 2.2e-113;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    323 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 382

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    383 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 442

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    443 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 502

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
             |||||||||||||||||||||||||||||||||||||||||||||

Db    503 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 547
```

RESULT 26
US-10-343-977-3
; Sequence 3, Application US/10343977
; Publication No. US20050112682A9
; GENERAL INFORMATION:
; APPLICANT: Kuhlemann, Rene
; APPLICANT: Koltermann, Andre
; APPLICANT: Kettling, Ulrich
; APPLICANT: Schwille, Petra
; TITLE OF INVENTION: Dual coloured fluorimetric protease assay
; FILE REFERENCE: 02342Sus6/JH/ml
; CURRENT APPLICATION NUMBER: US/10/343,977
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: PCT/EP01/09112
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 3
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Autofluorescent fusion protein derived from GFP variants of
; OTHER INFORMATION: e.g. Aequorea victoria and/or Discoma sp.
US-10-343-977-3

Query Match   100.0%; Score 1214; DB 5; Length 547;
Best Local Similarity   100.0%; Pred. No. 2.2e-113;
Matches 225; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    323 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 382

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    383 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 442

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
```

```
Db   443  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  502
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   503  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  547
```

RESULT 27
US-10-311-030-9
Sequence 9, Application US/10311030
Publication No. US20040171107A1
GENERAL INFORMATION:
APPLICANT: Nelson, David
APPLICANT: Zamiara, Elize
APPLICANT: Tsien, Roger
TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
FILE REFERENCE: 15916-032US1
CURRENT APPLICATION NUMBER: US/10/311,030
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: PCT/US01/04625
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: US 60/184,732
PRIOR FILING DATE: 2000-02-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 9
LENGTH: 226
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic construct
US-10-311-030-9

Query Match   99.7%; Score 1210; DB 4; Length 226;
Best Local Similarity 99.6%; Pred. No. 1.7e-113;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2  VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  61

Qy    61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  121

Qy   121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  181

Qy   181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   182  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  226
```

RESULT 28
US-10-152-296-2
Sequence 2, Application US/10152296
Publication No. US20030077730A1
GENERAL INFORMATION:
APPLICANT: Peelle, Beau
TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
FILE REFERENCE: 021044-0001100US
CURRENT APPLICATION NUMBER: US/10/152,296
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US 60/291,871
PRIOR FILING DATE: 2001-05-18
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 240
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:mammalian
OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
OTHER INFORMATION: "red" red fluorescent protein (RFP)
US-10-152-296-2

Query Match   99.7%; Score 1210; DB 4; Length 240;
Best Local Similarity 99.6%; Pred. No. 1.8e-113;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2  VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  61

Qy    61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  121

Qy   121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  181

Qy   181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   182  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  226
```

RESULT 29
US-10-739-656-2
Sequence 2, Application US/10739656
Publication No. US20040126850A1
GENERAL INFORMATION:
APPLICANT: Peelle, Beau
TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
FILE REFERENCE: 021044-0001100US
CURRENT APPLICATION NUMBER: US/10/739,656
CURRENT FILING DATE: 2003-12-17
PRIOR APPLICATION NUMBER: US/10/152,296
PRIOR FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US 60/291,871
PRIOR FILING DATE: 2001-05-18
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 240
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:mammalian
OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
OTHER INFORMATION: "red" red fluorescent protein (RFP)
US-10-739-656-2

Query Match   99.7%; Score 1210; DB 4; Length 240;
Best Local Similarity 99.6%; Pred. No. 1.8e-113;
Matches 224; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2  VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  61

Qy    61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  121

Qy   121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  181

Qy   181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   182  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYRTEGRHHLFL  226
```

RESULT 30
US-10-214-932-52
; Sequence 52, Application US/10214932
; Publication No. US20030100707A1
; GENERAL INFORMATION:
;   APPLICANT: HWANG, Inhwan
;   APPLICANT: KIM, Dae Heon
;   APPLICANT: LEE, Yong Jik
;   TITLE OF INVENTION: SYSTEM FOR DETECTING PROTEASE
;   FILE REFERENCE: APB02/US
;   CURRENT APPLICATION NUMBER: US/10/214,932
;   CURRENT FILING DATE: 2002-08-08
;   NUMBER OF SEQ ID NOS: 133
;   SOFTWARE: Patentin version 3.1
; SEQ ID NO 52
;   LENGTH: 545
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Synthetic Sequence
US-10-214-932-52

  Query Match   99.7%; Score 1210; DB 4; Length 545;
  Best Local Similarity   99.6%; Pred. No. 5.6e-113;
  Matches 224; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy   181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   182 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 31
US-10-214-932-76
; Sequence 76, Application US/10214932
; Publication No. US20030100707A1
; GENERAL INFORMATION:
;   APPLICANT: HWANG, Inhwan
;   APPLICANT: KIM, Dae Heon
;   APPLICANT: LEE, Yong Jik
;   TITLE OF INVENTION: SYSTEM FOR DETECTING PROTEASE
;   FILE REFERENCE: APB02/US
;   CURRENT APPLICATION NUMBER: US/10/214,932
;   CURRENT FILING DATE: 2002-08-08
;   NUMBER OF SEQ ID NOS: 133
;   SOFTWARE: Patentin version 3.1
; SEQ ID NO 76
;   LENGTH: 548
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Synthetic Sequence
US-10-214-932-76

  Query Match   99.7%; Score 1210; DB 4; Length 548;
  Best Local Similarity   99.6%; Pred. No. 5.7e-113;
  Matches 224; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 32
US-10-121-258-20
; Sequence 20, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: UC083.1CP2CP1
;   CURRENT APPLICATION NUMBER: US/10/121,258
;   CURRENT FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   NUMBER OF SEQ ID NOS: 78
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DsRed with I125R
US-10-121-258-20

  Query Match   99.4%; Score 1207; DB 4; Length 225;
  Best Local Similarity   99.6%; Pred. No. 3.3e-113;
  Matches 224; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 33
US-10-315-920-4
; Sequence 4, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
;   APPLICANT: Fradkov, Arcady Fedorovich
;   APPLICANT: Terskikh, Alexey
;   TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
;   TITLE OF INVENTION: FOR THEIR USE
;   FILE REFERENCE: CLON-077CIP
;   CURRENT APPLICATION NUMBER: US/10/315,920

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 VRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 61

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 121

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 181

Qy   181 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            |||||||||||||||||||||||||||||||||||||||||||||
Db   182 YMAKKPVQLPGYYYVVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 226

RESULT 33
US-10-315-920-4
; Sequence 4, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
;   APPLICANT: Fradkov, Arcady Fedorovich
;   APPLICANT: Terskikh, Alexey
;   TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
;   TITLE OF INVENTION: FOR THEIR USE
;   FILE REFERENCE: CLON-077CIP
;   CURRENT APPLICATION NUMBER: US/10/315,920

```
;   CURRENT FILING DATE: 2002-12-09
;   PRIOR APPLICATION NUMBER: 60/211,607
;   PRIOR FILING DATE: 2000-06-14
;   PRIOR APPLICATION NUMBER: PCT/US01/19097
;   PRIOR FILING DATE: 2001-06-13
;   NUMBER OF SEQ ID NOS: 22
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 4
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-4

Query Match          99.4%;  Score 1207;  DB 4;  Length 225;
Best Local Similarity  99.1%;  Pred. No. 3.3e-113;
Matches 223;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 34
US-10-931-304-20
; Sequence 20, Application US/10931304
; Publication No. US2005019768A1
; GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
;   PRIOR FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   NUMBER OF SEQ ID NOS: 110
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 20
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DsRed with I125R
US-10-931-304-20

Query Match          99.6%;  Score 1207;  DB 5;  Length 225;
Best Local Similarity  99.6%;  Pred. No. 3;
Matches 224;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225


RESULT 35
US-10-314-936-2
; Sequence 2, Application US/10314936
; Publication No. US2004011025A1
; GENERAL INFORMATION:
;   APPLICANT: Gibbs, Patrick D.L.
;   APPLICANT: Carter, Robert W.
;   APPLICANT: Schmale, Michael C.
;   TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
;   FILE REFERENCE: 638.004
;   CURRENT APPLICATION NUMBER: US/10/314,936
;   CURRENT FILING DATE: 2002-12-09
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 2
;   LENGTH: 236
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: mutant red fluorescent protein
US-10-314-936-2

Query Match          98.5%;  Score 1196;  DB 4;  Length 236;
Best Local Similarity  98.2%;  Pred. No. 4.6e-112;
Matches 221;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0;

Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||| ||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225


RESULT 36
US-10-314-936-4
; Sequence 4, Application US/10314936
; Publication No. US2004011025A1
; GENERAL INFORMATION:
;   APPLICANT: Gibbs, Patrick D.L.
;   APPLICANT: Carter, Robert W.
;   APPLICANT: Schmale, Michael C.
;   TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
;   FILE REFERENCE: 638.004
;   CURRENT APPLICATION NUMBER: US/10/314,936
;   CURRENT FILING DATE: 2002-12-09
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 4
```

```
;  LENGTH: 236
;  TYPE: PRT
;  ORGANISM: Artificial
;  FEATURE:
;  OTHER INFORMATION: mutant red fluorescent protein
US-10-314-936-4

Qy

Query Match             98.5%;  Score 1196;  DB 4;  Length 236;
Best Local Similarity   98.2%;  Pred. No. 4.6e-112;
Matches 221;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0;

Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            ||||||||||||||||||||:|||||||||||||||||||||||
Db    181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225


RESULT 37
US-11-021-014-2
;  Sequence 2, Application US/11021014
;  Publication No. US20050100954A1
;  GENERAL INFORMATION:
;  APPLICANT: Gibbs, Patrick D.L.
;  APPLICANT: Carter, Robert W.
;  APPLICANT: Schnale, Michael C.
;  TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
;  FILE REFERENCE: 638.004
;  CURRENT APPLICATION NUMBER: US/11/021,014
;  CURRENT FILING DATE: 2004-12-23
;  PRIOR APPLICATION NUMBER: US/10/314,936
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: Patentin version 3.1
US-11-021-014-2

;  SEQ ID NO 2
;  LENGTH: 236
;  TYPE: PRT
;  ORGANISM: Artificial
;  FEATURE:
;  OTHER INFORMATION: mutant red fluorescent protein

Query Match             98.5%;  Score 1196;  DB 6;  Length 236;
Best Local Similarity   98.2%;  Pred. No. 4.6e-112;
Matches 221;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0;

Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            ||||||||||||||||||||:|||||||||||||||||||||||
Db    181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225


RESULT 38
US-11-021-014-4
;  Sequence 4, Application US/11021014
;  Publication No. US20050100954A1
;  GENERAL INFORMATION:
;  APPLICANT: Gibbs, Patrick D.L.
;  APPLICANT: Carter, Robert W.
;  APPLICANT: Schnale, Michael C.
;  TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
;  FILE REFERENCE: 638.004
;  CURRENT APPLICATION NUMBER: US/11/021,014
;  CURRENT FILING DATE: 2004-12-23
;  PRIOR APPLICATION NUMBER: US/10/314,936
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: Patentin version 3.1
US-11-021-014-4

;  SEQ ID NO 4
;  LENGTH: 236
;  TYPE: PRT
;  ORGANISM: Artificial
;  FEATURE:
;  OTHER INFORMATION: mutant red fluorescent protein

Query Match             98.5%;  Score 1196;  DB 6;  Length 236;
Best Local Similarity   98.2%;  Pred. No. 4.6e-112;
Matches 221;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0;

Qy      1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKTI 180

Qy    181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
            ||||||||||||||||||||:|||||||||||||||||||||||
Db    181 YMAKKPVQLPGYYYVDSKLDITSHNKDYTIVEQYERTEGRHHLFL 225


RESULT 39
US-10-724-178-12
;  Sequence 12, Application US/10724178
;  Publication No. US20040137528A1
;  GENERAL INFORMATION:
;  APPLICANT: Odyssey Thera, Inc.
;  APPLICANT: Michnick, Stephen
;  APPLICANT: Macdonald, Marnie
;  APPLICANT: Lamerdin, Jane
;  TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
;  TITLE OF INVENTION: COMPLEMENTATION ASSAYS
;  FILE REFERENCE: ODDY007
;  CURRENT APPLICATION NUMBER: US/10/724,178
;  CURRENT FILING DATE: 2003-12-01
;  PRIOR APPLICATION NUMBER: US 60/461,133
;  PRIOR FILING DATE: 2003-04-09
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: Patentin version 3.0
;  SEQ ID NO 12
;  LENGTH: 226
;  TYPE: PRT
;  ORGANISM: Anthozoa discosoma
;  FEATURE:
;  NAME/KEY: misc feature
;  OTHER INFORMATION: Discosoma red ("dr") FP583
;  PUBLICATION INFORMATION:
;  AUTHORS: Matz, M.V. et al
;  TITLE: Fluorescent proteins from nonbioluminescent Anthozoa species
```

; JOURNAL: Nat. Biotech.
; VOLUME: 17
; ISSUE: 10
; PAGES: 969-73
; DATE: 1999-
; RELEVANT RESIDUES: (1)..(226)
US-10-724-178-12

  Query Match   98.5%; Score 1195.5; DB 4; Length 226;
  Best Local Similarity 99.1%; Pred. No. 4.9e-112;
  Matches 224; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Db  181 YMAKK-PVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        ||||| ||||||||||||||||||||||||||||||||||||||||
Qy  181 YMAKKAPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFLFL 226
```

RESULT 40
US-10-315-920-6
; Sequence 6, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Terskikh, Alexey
; APPLICANT: Fradkov, Arcady Fedorovich
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-6

  Query Match   98.1%; Score 1191; DB 4; Length 225;
  Best Local Similarity 97.8%; Pred. No. 1.4e-111;
  Matches 220; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Db    1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        | :|| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Db  181 YMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||| ||||||||||||||||||||||||||||
Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 41
US-10-442-148A-7
; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO
; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583_0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

  Query Match   97.3%; Score 1186; DB 4; Length 225;
  Best Local Similarity 97.3%; Pred. No. 4.4e-111;
  Matches 219; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Db    1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        | :|| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 42
US-10-442-148A-8
; Sequence 8, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO
; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583_0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 8
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-8

  Query Match   97.7%; Score 1186; DB 4; Length 239;

RESULT 43
US-10-121-258-4
; Sequence 4, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1"
US-10-121-258-4

Query Match 95.6%; Score 1160; DB 4; Length 225;
Best Local Similarity 96.0%; Pred. No. 1.8e-108;
Matches 216; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           | |::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||:|||| ||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGSFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 44
US-10-931-304-4
; Sequence 4, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger

RESULT 45
US-10-121-258-24
; Sequence 24, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1" with I125R mutation
US-10-121-258-24

Query Match 95.0%; Score 1153; DB 4; Length 225;

APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1"
US-10-931-304-4

Query Match 95.6%; Score 1160; DB 5; Length 225;
Best Local Similarity 96.0%; Pred. No. 1.8e-108;
Matches 216; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
           | |::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||:|||| ||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGSFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

Best Local Similarity 97.3%; Pred. No. 4.8e-111;
Matches 219; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          ||:| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKYYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKYYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQDSSLQDGSFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGETHKALKLKDGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 46
US-10-931-304-24

; Sequence 24, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1" with I125R mutation
US-10-931-304-24

Query Match   95.0%; Score 1153; DB 5; Length 225;
Best Local Similarity   95.6%; Conservative   3; Pred. No. 9.4e-108;
Matches 215; Conservative   3; Mismatches   7; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||:|||||||||||||||| |||||||||||||||||||| :|||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||||||||||||||||||||||||||||||||||||||| |||||
Db 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225

RESULT 47
US-10-423-688A-41

; Sequence 41, Application US/10423688A
; Publication No. US2004007848A1
; GENERAL INFORMATION:
; APPLICANT: Los Alamos National Laboratory
; APPLICANT: Waldo, Geoffrey S.
; TITLE OF INVENTION: Directed Evolution Methods for Improving Polypeptide Folding and
; TITLE OF INVENTION: Solubility and Superfolder Fluorescent Proteins Generated Thereby)
; FILE REFERENCE: S-100,608
; CURRENT APPLICATION NUMBER: US/10/423,688A
; CURRENT FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: 10/132,067
; PRIOR FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 41
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-423-688A-41

Query Match   93.2%; Score 1132; DB 4; Length 225;
Best Local Similarity   92.4%; Conservative   7; Pred. No. 1.2e-105;
Matches 208; Conservative   7; Mismatches 10; Indels   0; Gaps   0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |||: |||||||||||||| |||||||||||||||||||||| ||||||||||||||||
Db   1 MASSEDVIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQNVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||||| |||||||||||||| ||| ||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNFEDGGVVTVTQDSSLEDGCIIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          |||||||||||||||||||||||| |||||||||||||| ||||||||||||||| :|||
Db 121 KVKFIGVNFPSDGPVMQKKTMGWEPSTERLYPRDGVLKGDIHKALKLKDGGHYLVDIKSI 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||||||||| ||||||||||||| |||||||
Db 181 YMAKKPVQLPGYYYVDSKLDITNHNEDYTIVEQYERAEGRHHLFL 225

RESULT 48
US-10-121-258-6

; Sequence 6, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "dimer2"
US-10-121-258-6

Query Match   92.3%; Score 1121; DB 4; Length 226;
Best Local Similarity   92.8%; Conservative   7; Pred. No. 1.6e-104;
Matches 207; Conservative   7; Mismatches   9; Indels   0; Gaps   0;

Qy   3 SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS 62

RESULT 49
US-10-931-304-6
; Sequence 6, Application No. US20050196768A1
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "dimer2"
US-10-931-304-6

Query Match   92.3%; Score 1121; DB 5; Length 226;
Best Local Similarity   92.8%; Pred. No. 1.6e-104;
  Matches 207; Conservative   7; Mismatches   9; Indels   0; Gaps   0;

```
Qy    4 SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS 63
          ||||||||||||||||||:|||||||||||||||||||:: ||||||||||||||||||||
Db    4 SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS 63

Qy   64 PQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 PQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV 123

Qy  124 KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKSIYM 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db  124 KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKSIYM 183

Qy  183 AKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||:||||||||||||||||||||:|||||||
Db  184 AKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFL 226
```

RESULT 50
US-11-052-001-6
; Sequence 6, Application US/11052001
; Publication No. US20050233356A1
; GENERAL INFORMATION:
; APPLICANT: JONES, JOSHUA T.
; APPLICANT: HAHN, ANGELA TERESA
; APPLICANT: MEYER, TOBIAS
; TITLE OF INVENTION: LIVE-CELL BIOSENSOR POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: STAN-385
; CURRENT APPLICATION NUMBER: US/11/052,001
; CURRENT FILING DATE: 2005-02-03
; PRIOR APPLICATION NUMBER: 60/541,297
; PRIOR FILING DATE: 2004-02-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fluorescent fusion protein
US-11-052-001-6

Query Match   92.3%; Score 1121; DB 6; Length 602;
Best Local Similarity   92.8%; Pred. No. 6.1e-104;
  Matches 207; Conservative   7; Mismatches   9; Indels   0; Gaps   0;

```
Qy    3 SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS 62
          ||:|||||||||||||||:||||||||||||||||||||:::||||||||||||||||||
Db  142 SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS 201

Qy   63 PQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKV 122
          ||||||||:|||||||||||||||||||||||||||||||||||||||||||||::|||||
Db  202 PQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV 261

Qy  123 KFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYM 182
          ||:|:|||:|||||||||||||||||||||||||||||:||||||||||||||||||||||
Db  262 KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYM 321

Qy  183 AKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||||:|||||||||||||||||||:|||||||
Db  322 AKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFL 364
```

RESULT 51
US-10-006-922-44
; Sequence 44, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US2002019767 6A1e1 Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 225
; TYPE: PRT

```
;     ORGANISM: Artificial Sequence
;     FEATURE:
;     OTHER INFORMATION: hybrid construct
US-10-006-922-44

Query Match    92.8%;  Score 1119;  DB 4;  Length 225;
Best Local Similarity   91.6%;  Pred. No. 2.5e-104;
Matches  206;  Conservative   12;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||:||||||||||||||||||||||||||||:||||||||||||
Db    61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLKDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         :|||||||||||||||||||||||||||||::||||||||||||||||||| ||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 52
US-10-081-864-12
; Sequence 12, Application US/10081864
; Publication No. US20030022287A1
;     GENERAL INFORMATION:
;     APPLICANT: Lukyanov, Sergey
;     APPLICANT: Lukyanov, Konstantin
;     APPLICANT: Yanushevich, Yuriy
;     APPLICANT: Savitsky, Alexandr
;     APPLICANT: Fradkov, Arcady
;     TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
;     TITLE OF INVENTION: Methods for Using the Same
;     FILE REFERENCE: CLON-067
;     CURRENT APPLICATION NUMBER: US/10/081,864
;     CURRENT FILING DATE: 2002-06-19
;     PRIOR APPLICATION NUMBER: 10/006,922
;     PRIOR FILING DATE: 2001-12-04
;     PRIOR APPLICATION NUMBER: 60/270,983
;     PRIOR FILING DATE: 2001-02-21
;     NUMBER OF SEQ ID NOS: 30
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 12
;     LENGTH: 225
;     TYPE: PRT
;     ORGANISM: Artificial Sequence
;     FEATURE:
;     OTHER INFORMATION: hybrid protein
US-10-081-864-12

Query Match    92.8%;  Score 1119;  DB 4;  Length 225;
Best Local Similarity   91.6%;  Pred. No. 2.5e-104;
Matches  206;  Conservative   12;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDI 60

Qy    61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||:||||||||||||||||||||||||||||:||||||||||||
Db    61 LSPQFQYGSKYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLKDGCFIY 120

Qy   121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         :|||||||||||||||||||||||||||||::||||||||||||||||||| ||||||
Db   121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

```
Db   181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225

RESULT 53
US-10-931-304-81
; Sequence 81, Application US/10931304
; Publication No. US20050196768A1
;     GENERAL INFORMATION:
;     APPLICANT: Tsien, Roger
;     APPLICANT: Campbell, Robert
;     APPLICANT: Baird, Geoffrey
;     TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;     TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;     FILE REFERENCE: 39754-0831CP2CP3
;     CURRENT APPLICATION NUMBER: US/10/931,304
;     CURRENT FILING DATE: 2004-08-30
;     PRIOR APPLICATION NUMBER: 10/209,208
;     PRIOR FILING DATE: 2002-07-29
;     PRIOR APPLICATION NUMBER: 10/121,258
;     PRIOR FILING DATE: 2002-04-10
;     PRIOR APPLICATION NUMBER: 09/866,538
;     PRIOR FILING DATE: 2001-05-24
;     PRIOR APPLICATION NUMBER: 09/794,308
;     PRIOR FILING DATE: 2001-02-26
;     NUMBER OF SEQ ID NOS: 110
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 81
;     LENGTH: 234
;     TYPE: PRT
;     ORGANISM: Artificial Sequence
;     FEATURE:
;     OTHER INFORMATION: DsRed polypeptide variant "dimer2.2MM (dimer3) (dtomato)
US-10-931-304-81

Query Match    90.2%;  Score 1095;  DB 5;  Length 234;
Best Local Similarity   91.0%;  Pred. No. 7e-102;
Matches  201;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     5 KNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQ 64
         ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db     6 EEVIKEFMRFKVRMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ 65

Qy    65 FQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF 124
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66 FMYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKVKM 125

Qy   125 IGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 184
         |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db   126 RGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYMAK 185

Qy   185 KPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         ||||||||||||||||||||||||||||||||||||:|||||
Db   186 KPVQLPGYYYVDITKLDITSHNEDYTIVEQYERSEGRHHLFL 226
```

```
RESULT 54
US-10-931-304-106
; Sequence 106, Application US/10931304
; Publication No. US20050196768A1
;     GENERAL INFORMATION:
;     APPLICANT: Tsien, Roger
;     APPLICANT: Campbell, Robert
;     APPLICANT: Baird, Geoffrey
;     TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;     TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;     FILE REFERENCE: 39754-0831CP2CP3
;     CURRENT APPLICATION NUMBER: US/10/931,304
;     CURRENT FILING DATE: 2004-08-30
;     PRIOR APPLICATION NUMBER: 10/209,208
;     PRIOR FILING DATE: 2002-07-29
;     PRIOR APPLICATION NUMBER: 10/121,258
;     PRIOR FILING DATE: 2002-04-10
;     PRIOR APPLICATION NUMBER: 09/866,538
```

; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 106
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "tdTomato"
US-10-931-304-106

Query Match   90.2%; Score 1095; DB 5; Length 476;
Best Local Similarity 91.0%; Pred. No. 1.9e-101;
Matches 201; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

```
Qy    5 KNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQ 64
          ||||||||||||||||::|||||||||||||||:||||||||||||||||||||||||||
Db    6 EEVIKEFMRFKVRMEGSMNGHEFEIEGEGEGRPYEGTQTATAKLKVTKGGPLPFAWDILSPQ 65

Qy   65 FQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKF 124
          ||||||:|||||||||||||||||||||||||:|||||:|||||||||||||||||:|||:|
Db   66 FMYGSKAYVKHPADIPDYKKLSFPEGFKWERVNFEDGGLVTVTQDSSLQDGTLIYKVKM 125

Qy  125 IGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAK 184
          ||:||||||||||||||||||||||||||||||:|||:||||||||||||||:|||||||
Db  126 RGTNFPPDGPVWQKKTMGWEASTERLYPRDGVLKGDIHQALKLKDGGHYLVEFKTIYMAK 185

Qy  185 KPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||:|||||||||:||||||||||||||||||:|||:||||
Db  186 KPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFL 226
```

RESULT 55
US-10-006-922-18
; Sequence 18, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 18
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Discosoma species

Query Match   89.4%; Score 1085.5; DB 4; Length 230;
Best Local Similarity 88.1%; Pred. No. 6.2e-101;
Matches 199; Conservative 17; Mismatches 9; Indels 1; Gaps 1;

RESULT 56
US-10-161-403-40
; Sequence 40, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 40
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-161-403-40

Query Match   89.4%; Score 1085.5; DB 4; Length 230;
Best Local Similarity 88.1%; Pred. No. 6.2e-101;
Matches 199; Conservative 17; Mismatches 9; Indels 1; Gaps 1;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          |::|||||||||||||||||::||||||:|||||||||||||::|||:||||||||||:||
Db    1 MSCSKNVIKEFMRFKVRMEGSMNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAPDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
          |||||||||||||||||||||||||||||||||||:||||||||||||||||||||:|||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVNFEDGGVVTVSQDSSLKDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
          :|||||||||||||||::|:|||||:|||||||||||||:|||:||:|||||||||||||
Db  121 EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI 180

Qy  181 YMAKKP-VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          ||.|||  |||||||||||||:|||||||||||||||::|||||:|:
Db  181 YMVKKPSVQLPGYYYVDSKLDMTSHNEDYTIVEQYEKTQGRHHPFI 226
```

RESULT 57
US-11-006-076-40
; Sequence 40, Application US/11006076
; Publication No. US20050181506A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward

```
;    APPLICANT:  Perez, Carl
;    APPLICANT:  Lindenbaum, Michael
;    APPLICANT:  Greene, Amy
;    APPLICANT:  Leung, Josephine
;    APPLICANT:  Fleming, Elena
;    APPLICANT:  Stewart, Sandra
;    APPLICANT:  Shellard, Joan
;    TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
;    FILE REFERENCE: 17084-022002/420B
;    CURRENT APPLICATION NUMBER: US/11/006,076
;    CURRENT FILING DATE: 2004-12-06
;    PRIOR APPLICATION NUMBER: 60/294,758
;    PRIOR FILING DATE: 2001-05-30
;    PRIOR APPLICATION NUMBER: 60/366,891
;    PRIOR FILING DATE: 2002-03-21
;    NUMBER OF SEQ ID NOS: 129
;    SOFTWARE: FastSEQ for Windows Version 4.0
;
;    SEQ ID NO 40
;    LENGTH: 230
;    TYPE: PRT
;    ORGANISM: Discosoma species
US-11-006-076-40

Query Match      89.4%;  Score 1085.5;  DB 6;  Length 230;
Best Local Similarity   88.1%;  Pred. No. 6.2e-101;
Matches 199;  Conservative  17;  Mismatches  9;  Indels  1;  Gaps  1;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         |||||||||||||||||||||||||||||||||||||||||       ||||||||| ||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHCSVSKLMVTKGGPLPFAFDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         :||||||||||||||||::: ||||:||||||||||||| || |:| | |||||||||||
Db  121  EVKFIGVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSI  180

Qy  181  YMAKKP--VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         ||.||    ||||||||||||:|||||||||:||||||:| ||| |.
Db  181  YMVKKPSVQLPGYYYVDSKLDMTSHNEDYTVVEQYEKTQGRHPFI  226


RESULT 58
US-10-006-922-46
;    Sequence 46, Application US/10006922
;    Publication No. US20020197676A1
;    GENERAL INFORMATION:
;    APPLICANT:  Lukyanov, Sergey A
;    APPLICANT:  Fradkov, Arcady F.
;    APPLICANT:  Labas, Yulii A.
;    APPLICANT:  Matz, Mikhail V.
;    APPLICANT:  Terskikh, Alexey
;    TITLE OF INVENTION: No. US20020197676A1e1 Chromophores/Fluorophores and
;    TITLE OF INVENTION: Methods for Using the Same
;    FILE REFERENCE: CLON-035CIP
;    CURRENT APPLICATION NUMBER: US/10/006,922
;    CURRENT FILING DATE: 2001-12-04
;    PRIOR APPLICATION NUMBER: 09/120,330
;    PRIOR FILING DATE: 1998-12-11
;    PRIOR APPLICATION NUMBER: 09/457,898
;    PRIOR FILING DATE: 1999-12-09
;    PRIOR APPLICATION NUMBER: 09/458,144
;    PRIOR FILING DATE: 1999-12-09
;    PRIOR APPLICATION NUMBER: 09/458,477
;    PRIOR FILING DATE: 1999-12-09
;    PRIOR APPLICATION NUMBER: 09/457,556
;    PRIOR FILING DATE: 1999-12-09
;    PRIOR APPLICATION NUMBER: 09/444,338
;    PRIOR FILING DATE: 1999-11-19
;    NUMBER OF SEQ ID NOS: 46
;    SOFTWARE: FastSEQ for Windows Version 4.0
```

```
;    SEQ ID NO 46
;    LENGTH: 205
;    TYPE: PRT
;    ORGANISM: Discosoma species
US-10-006-922-46

Query Match      86.2%;  Score 1046;  DB 4;  Length 205;
Best Local Similarity   87.8%;  Pred. No. 5.1e-97;
Matches 197;  Conservative  3;  Mismatches  5;  Indels  20;  Gaps  1;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
             :|||||||||||||||||||||||  |                  |||||||||||
Db  121  KSSSLALNFPSDGPVMQKKTMGWEASTERL------------GHYLVEFKSI  160

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL  225
         |||||||||||||||||||||||||||||||||||:|||||||||
Db  161  IMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERSEGRHHLFL  205


RESULT 59
US-10-121-258-8
;    Sequence 8, Application US/10121258
;    Publication No. US20030059835A1
;    GENERAL INFORMATION:
;    APPLICANT:  Tsien, Roger
;    APPLICANT:  Campbell, Robert
;    TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;    TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;    FILE REFERENCE: UC083.1CP2CP1
;    CURRENT APPLICATION NUMBER: US/10/121,258
;    CURRENT FILING DATE: 2002-04-10
;    PRIOR APPLICATION NUMBER: 09/794,308
;    PRIOR FILING DATE: 2001-02-26
;    PRIOR APPLICATION NUMBER: 09/866,538
;    PRIOR FILING DATE: 2001-05-24
;    NUMBER OF SEQ ID NOS: 78
;    SOFTWARE: FastSEQ for Windows Version 4.0
;    SEQ ID NO 8
;    LENGTH: 225
;    TYPE: PRT
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: DsRed polypeptide variant "mRFP1"
US-10-121-258-8

Query Match      84.1%;  Score 1021;  DB 4;  Length 225;
Best Local Similarity   86.9%;  Pred. No. 1.9e-94;
Matches 192;  Conservative  5;  Mismatches  24;  Indels  0;  Gaps  0;

Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI   60
         |::||||||||||||||||||:||||||||||||||||||   :||||||||||||||||
Db    1  MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI   60

Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY  120
         ||||||||||.||||||||||.|||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY  120

Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI  180
         |||  |:|||||||||||||||||||||:| |:||:|||||:| ||:|||||   |  ::
Db  121  KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT  180

Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRH  221
         |||||||||||||||||:|||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDTIKLDITSHNEDYTIVEQYERAEGRH  221
```

RESULT 60
US-10-724-178-16
; Sequence 16, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: MacDonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 16
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: "mRFP1", an engineered monomeric form of DsRed FP
US-10-724-178-16

Query Match 84.1%; Score 1021; DB 4; Length 225;
Best Local Similarity 86.9%; Pred. No. 1.9e-94;
Matches 192; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         |::|||||||||||||||||||:||||||||||||||||:|:|:||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI  60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||:|||||||||||:||||||||||||||||||||||||||||||||||:|||
Db  61 LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||:.|.||||||||||||||||||||||:||.||.||||:.|.||||||||| :|.:.
Db 121 KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 221
         |||||||||||||||||||||||||||||||||||||:||||
Db 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERTEGRH 221
```

RESULT 61
US-10-931-304-8
; Sequence 8, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "mRFP1"
US-10-931-304-8

Query Match 84.1%; Score 1021; DB 5; Length 225;
Best Local Similarity 86.9%; Pred. No. 1.9e-94;
Matches 192; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         |::|||||||||||||||||:|||||||||||||||||:|:|:||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI  60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||:|||||||||||:||||||||||||||||||||||||||||||||||:|||
Db  61 LSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||:.|.||||||||||||||||||||||:||.||.||||:.|.||||||||| :|.:.
Db 121 KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH 221
         |||||||||||||||||||||||||||||||||||||:||||
Db 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERTEGRH 221
```

RESULT 62
US-10-931-304-108
; Sequence 108, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "mGrape1"
US-10-931-304-108

Query Match 83.5%; Score 1014; DB 5; Length 225;
Best Local Similarity 86.4%; Pred. No. 9.8e-94;
Matches 191; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

```
Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI  60
         |::|||||||||||||||||:|||||||||||||||||:|:|:||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI  60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||:|||:|||||||||||:||||||||||||||||||||||||||||||||||:|||
Db  61 LSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||:.|.||||||||||||||||||||||||||.||.||||:.|.||||||||| :|.:.
Db 121 KVKLRGTNFPSDGPVMQKKTMGWEASTERLYPEDGALKGEIKMRLKLKDGGHYDABAKTT 180
```

RESULT 63
US-10-931-304-79
Sequence 79, Application US/10931304
Publication No. US2005019676841
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
APPLICANT: Baird, Geoffrey
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: 39754-0831CP2CP3
CURRENT APPLICATION NUMBER: US/10/931,304
CURRENT FILING DATE: 2004-08-30
PRIOR APPLICATION NUMBER: 10/209,208
PRIOR FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 79
LENGTH: 225
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: DsRed polypeptide variant "mRFP1.1"
US-10-931-304-79

Query Match   83.4%; Score 1012; DB 5; Length 225;
Best Local Similarity 86.0%; Pred. No. 1.6e-93;
Matches 190; Conservative 6; Mismatches 25; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          | |:||||||||||||||||| ||||||||||||||||||  | :|||||||||||||||
Db    1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFMYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||: |.|||||||||||||||||||:|:||| |:||||||| |:|||||||||| : |
Db  121 KVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         ||||||||||| | | :||||||||||||||||||| |||||
Db  181 YMAKKPVQLPGAYKDIKLDITSHNEDYTIVEQYERAEGRH 221
```

RESULT 64
US-10-931-304-104
Sequence 104, Application US/10931304
Publication No. US2005019676841
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
APPLICANT: Baird, Geoffrey
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: 39754-0831CP2CP3
CURRENT APPLICATION NUMBER: US/10/931,304
CURRENT FILING DATE: 2004-08-30
PRIOR APPLICATION NUMBER: 10/209,208
PRIOR FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 104
LENGTH: 225
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: DsRed polypeptide variant "mHoneydew"
US-10-931-304-104

Query Match   82.9%; Score 1007; DB 5; Length 225;
Best Local Similarity 85.1%; Pred. No. 5e-93;
Matches 188; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
          | |:|||||||||||||||| ||||||||||||||||||  | :|||||||||||||||
Db    1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKYTVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||:.|||: ||||||||||:|||||||||||||||||||||||||||||||||||||
Db   61 LSPQFMWGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||: |.||||||||||||||||  :|||:||| |:|||||| |:||||||||| : |
Db  121 KVKLRGTNFPSDGPVMQKKTMGWAATTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         ||||||||||| | | :||||||||||||||||||| |||||
Db  181 YMAKKPVQLPGAYKDGKLDITSHNEDYTIVEQYERAEGRH 221
```

RESULT 65
US-10-931-304-100
Sequence 100, Application US/10931304
Publication No. US2005019676841
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
APPLICANT: Baird, Geoffrey
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: 39754-0831CP2CP3
CURRENT APPLICATION NUMBER: US/10/931,304
CURRENT FILING DATE: 2004-08-30
PRIOR APPLICATION NUMBER: 10/209,208
PRIOR FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 100
LENGTH: 225
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: DsRed polypeptide variant "mTangerine"
US-10-931-304-100

Query Match   82.5%; Score 1002; DB 5; Length 225;
Best Local Similarity 85.5%; Pred. No. 1.6e-92;
Matches 189; Conservative 6; Mismatches 26; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
```

RESULT 66
US-10-931-304-85
; Sequence 85, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
;   PRIOR FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   NUMBER OF SEQ ID NOS: 110
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 85
;   LENGTH: 236
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DsRed polypeptide variant "OrS4-9"
US-10-931-304-85

Query Match   81.9%; Score 994; DB 5; Length 236;
Best Local Similarity 86.5%; Pred. No. 1.1e-91;
Matches 186; Conservative   5; Mismatches 24; Indels   0; Gaps   0;

```
Db    1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI  60
          ::|||||||||||||.||||||||||||||||||||   |.||||||||||||||||
Qy
Qy   61 LSPQFCYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||.||||||||||.||||||||||||||||||||||||||||||||.|||
Db   61 LSPQFCYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120
Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||  |.|||||||||||||||||||.||.||.||.||||| |  |||||||||  |. |
Db  121 KVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180
Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         |||||||||||  | | |||||||||||||| ||| ||||
Db  181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIVELYERAEGRH 221
```

RESULT 67
US-10-931-304-89
; Sequence 89, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
;   PRIOR FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   NUMBER OF SEQ ID NOS: 110
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 89
;   LENGTH: 236
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DsRed polypeptide variant "mRFP (F2Q6) (mGrape2)"
US-10-931-304-89

Query Match   81.7%; Score 992; DB 5; Length 236;
Best Local Similarity 86.0%; Pred. No. 1.7e-91;
Matches 185; Conservative   6; Mismatches 24; Indels   0; Gaps   0;

```
Qy    7 VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ  66
         :|||||||||||||:||||||||||||||||||||   |.||||||||||||||||||
Db   12 IIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFT  71
Qy   67 YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG 126
         ||||.||||||||||.||||||||||||||||||||||||||||||||||.||||||.|
Db   72 YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRG 131
Qy  127 VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP 186
         .|||||||||||||||||||.||.||.||.|||||  | |||||||||.|  |. |||||
Db  132 TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKP 191
Qy  187 VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         |||||  | | |||||||||||||| ||| ||||
Db  192 VQLPGAYKLDYKLDITSHNEDYTIVEQYERAEGRH 226
```

RESULT 68
US-10-931-304-102
; Sequence 102, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
;   PRIOR FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   NUMBER OF SEQ ID NOS: 110
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 102
;   LENGTH: 236
;   TYPE: PRT

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Dsred polypeptide variant "mOrange (mOFP1)"
US-10-931-304-102

Qy  Query Match
    Best Local Similarity  81.5%;   Score 989;  DB 5;   Length 236;
    Matches 185;  Conservative 5;  Mismatches 25;  Indels 0;  Gaps 0;

Qy    7  VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ  66
         ||||||||||:|||:|||||||||||||||||||| ||:|||||||||||||||||||
Db   12  IIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFT  71

Qy   67  YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG  126
         ||||.||||||||||| |||||||||||||||||||||||||||||||||||:||||| |
Db   72  YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRG  131

Qy  127  VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP  186
         .|||||||||||||||||||:|||:||||||| ||||:|||||||| :|: |||||
Db  132  TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYTSEVKTTYKAKKP  191

Qy  187  VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH  221
         |||||.|:..:||||||||||||||||||||.||||
Db  192  VQLPGAYIVGIKLDITSHNEDYTIVEQYERAEGRH  226


RESULT 69
US-10-931-304-83
Sequence 83, Application US/10931304
;   Publication No. US2005019768A1
;   GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
;   PRIOR FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   NUMBER OF SEQ ID NOS: 110
;   SEQ ID NO 83
;   LENGTH: 236
;   TYPE: PRT
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Dsred polypeptide variant "mRFP1.5"
US-10-931-304-83

Qy  Query Match
    Best Local Similarity  85.6%;   Score 988;  DB 5;   Length 236;
    Matches 184;  Conservative 6;  Mismatches 25;  Indels 0;  Gaps 0;

Qy    7  VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ  66
         ||||||||||:|||:|||||||||||||||||||| ||:|||||||||||||||||||
Db   12  IIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFM  71

Qy   67  YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG  126
         ||||.||||||||||| |||||||||||||||||||||||||||||||||||:||||| |
Db   72  YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRG  131

Qy  127  VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP  186
         .||||||||||||||||||||:|||:||||||| ||||:|||||||| :|: |||||
Db  132  TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKP  191

Qy  187  VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH  221
         |||||.|:..:||||||||||||||||||||.||||
Db  192  VQLPGAYNVIKLDITSHNEDYTIVEQYERAEGRH  226
```

```
Qy  187  VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH  221
         |||||.|: ..:||||||||||||||||||.||||
Db  192  VQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRH  226


RESULT 70
US-10-931-304-92
Sequence 92, Application US/10931304
;   Publication No. US2005019768A1
;   GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
;   PRIOR FILING DATE: 2002-04-10
;   PRIOR APPLICATION NUMBER: 09/866,538
;   PRIOR FILING DATE: 2001-05-24
;   PRIOR APPLICATION NUMBER: 09/794,308
;   PRIOR FILING DATE: 2001-02-26
;   NUMBER OF SEQ ID NOS: 110
;   SEQ ID NO 92
;   LENGTH: 236
;   TYPE: PRT
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Dsred polypeptide variant "mRFP2 (mcherry)"
US-10-931-304-92

Qy  Query Match
    Best Local Similarity  81.3%;   Score 987;  DB 5;   Length 236;
    Matches 184;  Conservative 7;  Mismatches 24;  Indels 0;  Gaps 0;

Qy    7  VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ  66
         ||||||||||:|||:|||||||||||||||||||| ||:|||||||||||||||||||
Db   12  IIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFM  71

Qy   67  YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG  126
         ||||.||||||||||| |||||||||||||||||||||||||||||||||||:||||| |
Db   72  YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRG  131

Qy  127  VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP  186
         .||||||||||||||||||||:|||:||||||| ||||:|||||||| :|: |||||
Db  132  TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKP  191

Qy  187  VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH  221
         |||||.|:..:||||||||||||||||||||.||||
Db  192  VQLPGAYNVIKLDITSHNEDYTIVEQYERAEGRH  226


RESULT 71
US-10-931-304-94
Sequence 94, Application US/10931304
;   Publication No. US2005019768A1
;   GENERAL INFORMATION:
;   APPLICANT: Tsien, Roger
;   APPLICANT: Campbell, Robert
;   APPLICANT: Baird, Geoffrey
;   TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;   FILE REFERENCE: 39754-0831CP2CP3
;   CURRENT APPLICATION NUMBER: US/10/931,304
;   CURRENT FILING DATE: 2004-08-30
;   PRIOR APPLICATION NUMBER: 10/209,208
;   PRIOR FILING DATE: 2002-07-29
;   PRIOR APPLICATION NUMBER: 10/121,258
```

```
            PRIOR FILING DATE: 2002-04-10
            PRIOR APPLICATION NUMBER: 09/866,538
            PRIOR FILING DATE: 2001-05-24
            PRIOR APPLICATION NUMBER: 09/794,308
            PRIOR FILING DATE: 2001-02-26
            NUMBER OF SEQ ID NOS: 110
            SOFTWARE: FastSEQ for Windows Version 4.0
            SEQ ID NO 94
            LENGTH: 235
            TYPE: PRT
            ORGANISM: Artificial Sequence
            FEATURE:
            OTHER INFORMATION: DsRed polypeptide variant "mYOFP (74-11)"
            FEATURE:
            NAME/KEY: VARIANT
            LOCATION: 62
            OTHER INFORMATION: Xaa = Any Amino Acid
          US-10-931-304-94

          Query Match      81.2%;   Score 986;   DB 5;   Length 235;
          Best Local Similarity   85.6%;   Pred. No. 7e-91;
          Matches 184;   Conservative 5;   Mismatches 26;   Indels 0;   Gaps 0;

          Qy      7 VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ 66
                    :|||||||||||||| ||||||||||||||||||||   ||||||||||||||||| | |
          Db     12 IIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGFQTAKLKVTKGGPLPFAWDILTPNFT 71

          Qy     67 YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG 126
                    |||| |||||||||| ||||||||||||||||||||||||||||||||||| |||||| |
          Db     72 YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRG 131

          Qy    127 VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP 186
                    |||||||||||||||||||| |  :|  || ||||  |  |||||||  : | | |||||
          Db    132 TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYTSEVKTYKAKKP 191

          Qy    187 VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
                    |||||   | |||||||||||||||||| | ||||
          Db    192 VQLPGAYIVGIKLDITSHNEDYTIVEQYERAEGRH 226


          RESULT 72
          US-10-931-304-96
          Sequence 96, Application US/10931304
          Publication No. US20050196768A1
          GENERAL INFORMATION:
            APPLICANT: Tsien, Roger
            APPLICANT: Campbell, Robert
            APPLICANT: Baird, Geoffrey
            TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
               PROTEIN VARIANTS AND METHODS FOR MAKING SAME
            FILE REFERENCE: 39754-0831CP2CP3
            CURRENT APPLICATION NUMBER: US/10/931,304
            CURRENT FILING DATE: 2004-08-30
            PRIOR APPLICATION NUMBER: 10/209,208
            PRIOR FILING DATE: 2002-07-29
            PRIOR APPLICATION NUMBER: 10/121,258
            PRIOR FILING DATE: 2002-04-10
            PRIOR APPLICATION NUMBER: 09/866,538
            PRIOR FILING DATE: 2001-05-24
            PRIOR APPLICATION NUMBER: 09/794,308
            PRIOR FILING DATE: 2001-02-26
            NUMBER OF SEQ ID NOS: 110
            SOFTWARE: FastSEQ for Windows Version 4.0
            SEQ ID NO 96
            LENGTH: 236
            TYPE: PRT
            ORGANISM: Artificial Sequence
            FEATURE:
            OTHER INFORMATION: DsRed polypeptide variant "mROFP (A2/6-6)"
          US-10-931-304-96

          Query Match      80.1%;   Score 973;   DB 5;   Length 236;
          Best Local Similarity   84.7%;   Pred. No. 1.4e-89;
```

```
          Matches 182;   Conservative 6;   Mismatches 27;   Indels 0;   Gaps 0;

          Qy      7 VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ 66
                    :|||||||||||||| ||||||||||||||||||||   ||||||||||||||||| | |
          Db     12 IIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGFQTAKLKVTKGGPLPFAWDILTPNFT 71

          Qy     67 YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG 126
                    |||| |||||||||| ||||||||||||||||||||||||||||||||||| |||||| |
          Db     72 YGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRG 131

          Qy    127 VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP 186
                    |||||||||||||||||||| |  :|  || ||||  |  |||||||  : | | |||||
          Db    132 TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTYKAKKP 191

          Qy    187 VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
                    |||||   | |||||||||||||||||:| |||||
          Db    192 VQLPGAYIVGIKLDITSHNEDYTIVELYERAEGRH 226


          RESULT 73
          US-10-931-304-98
          Sequence 98, Application US/10931304
          Publication No. US20050196768A1
          GENERAL INFORMATION:
            APPLICANT: Tsien, Roger
            APPLICANT: Campbell, Robert
            APPLICANT: Baird, Geoffrey
            TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
               PROTEIN VARIANTS AND METHODS FOR MAKING SAME
            FILE REFERENCE: 39754-0831CP2CP3
            CURRENT APPLICATION NUMBER: US/10/931,304
            CURRENT FILING DATE: 2004-08-30
            PRIOR APPLICATION NUMBER: 10/209,208
            PRIOR FILING DATE: 2002-07-29
            PRIOR APPLICATION NUMBER: 10/121,258
            PRIOR FILING DATE: 2002-04-10
            PRIOR APPLICATION NUMBER: 09/866,538
            PRIOR FILING DATE: 2001-05-24
            PRIOR APPLICATION NUMBER: 09/794,308
            PRIOR FILING DATE: 2001-02-26
            NUMBER OF SEQ ID NOS: 110
            SOFTWARE: FastSEQ for Windows Version 4.0
            SEQ ID NO 98
            LENGTH: 236
            TYPE: PRT
            ORGANISM: Artificial Sequence
            FEATURE:
            OTHER INFORMATION: DsRed polypeptide variant "mStrawberry"
          US-10-931-304-98

          Query Match      80.1%;   Score 973;   DB 5;   Length 236;
          Best Local Similarity   84.7%;   Pred. No. 1.4e-89;


          RESULT 74
          US-10-931-304-87
```

RESULT 75
US-10-724-178-1041
Sequence 1041, Application US/10724178
Publication No. US20040137528A1
GENERAL INFORMATION:
APPLICANT: Odyssey Thera, Inc.
APPLICANT: Michnick, Stephen
APPLICANT: MacDonald, Marnie
APPLICANT: Lamerdin, Jane
TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
TITLE OF INVENTION: COMPLEMENTATION ASSAYS
FILE REFERENCE: ODDY007
CURRENT APPLICATION NUMBER: US/10/724,178
CURRENT FILING DATE: 2003-12-01
PRIOR APPLICATION NUMBER: US 60/461,133
PRIOR FILING DATE: 2003-04-09
NUMBER OF SEQ ID NOS: 1067
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1041
LENGTH: 183
TYPE: PRT
ORGANISM: Artificial
FEATURE:

Sequence 87, Application US/10931304
Publication No. US20050196768A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger
APPLICANT: Campbell, Robert
APPLICANT: Baird, Geoffrey
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: 39754-0B3CP2CP3
CURRENT APPLICATION NUMBER: US/10/931,304
CURRENT FILING DATE: 2004-08-30
PRIOR APPLICATION NUMBER: 10/209,208
PRIOR FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: 10/121,258
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/866,538
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 87
TYPE: PRT
LENGTH: 236
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: DsRed polypeptide variant "Y1.3 (myOFP 1.3) (mBanana)"
US-10-931-304-87

```
Query Match          79.1%;  Score 960;  DB 5;  Length 236;
Best Local Similarity 83.7%;  Pred. No. 2.9e-88;
Matches 180;  Conservative 5;  Mismatches 30;  Indels 0;  Gaps 0;

Qy    7 VIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQFQ 66
          |||||| ||||||  |||||||||||||||||||||       |||||||||||||||||||
Db   12 VIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFC 71

Qy   67 YGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFIG 126
          ||||| |||| |||||| ||||||||||||| |||||||||| |||||||||| ||||||
Db   72 YGSKAYVKHPTGIPDYFKLSFPEGFKWERVWNFEDGGVVTVAQDSSLQDGEFIYKVKLRG 131

Qy  127 VNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKKP 186
         |||||||||||||||||| |||| |||||||||||    |||||||||:  |||||||
Db  132 TNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYSAETKTTYKAKKP 191

Qy  187 VQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         |||||  |   ||||||||||||||| ||||||
Db  192 VQLPGAYIAGEKIDITSHNEDYTIVELYERAEGRH 226
```

; OTHER INFORMATION: modif. frag.; RFP F1F, with position 1 Met removed
US-10-724-178-1041

```
Query Match          69.9%;  Score 848;  DB 4;  Length 183;
Best Local Similarity 87.4%;  Pred. No. 4.1e-77;
Matches 159;  Conservative 5;  Mismatches 18;  Indels 0;  Gaps 0;

Qy    3 SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS 62
          ||:|||||||||||||||  |||||||||||||||||||       ||||||||||||||
Db    2 SSBDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS 61

Qy   63 PQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKV 122
         ||||||||| |||||||||| |||||||||||| |||||||||||||||||||| ||||
Db   62 PQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVWNFEDGGVVTVTQDSSLQDGEFIYKV 121

Qy  123 KFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYM 182
         |  |||||||||||||||||| |||| |||||||||||    |||||||||:  |||
Db  122 KLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYM 181

Qy  183 KK 184
         ||
Db  182 AK 183
```

Search completed: January 11, 2006, 02:12:06
Job time : 64 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 11, 2006, 02:08:34 ; Search time 9 Seconds
(without alignments)
212.341 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG.........EDYTIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 61141 seqs, 8493638 residues
Total number of hits satisfying chosen parameters: 7

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 60%
Maximum Match 100%
Listing first 500 summaries

Database : Published Applications AA_New:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	225	6	US-10-209-208-1	Sequence 1, Appli
2	1207	99.4	225	6	US-10-209-208-20	Sequence 20, Appl
3	1160	95.6	225	6	US-10-209-208-4	Sequence 4, Appl
4	1153	95.0	225	6	US-10-209-208-24	Sequence 24, Appl
5	1121	92.3	226	6	US-10-209-208-6	Sequence 6, Appl
6	1021	84.1	225	6	US-10-209-208-8	Sequence 8, Appl
7	1012	83.4	225	6	US-10-209-208-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-10-209-208-1
; Sequence 1, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; OTHER INFORMATION: wild-type DsRed
US-10-209-208-1
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(225)

Query Match 100.0%; Score 1214; DB 6; Length 225;
Best Local Similarity 100.0%; Pred. No. 2.3e-110;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60

Qy   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120

Qy  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180

Qy  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
        |||||||||||||||||||||||||||||||||||||||||||||
Db  181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

RESULT 2
US-10-209-208-20
; Sequence 20, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
US-10-209-208-20
; SEQ ID NO 20
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed with I125R

Query Match 99.4%; Score 1207; DB 6; Length 225;
Best Local Similarity 99.6%; Pred. No. 1.1e-109;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
```

RESULT 3
US-10-209-208--4

```
Db    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
```

;
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1"
US-10-209-208-4

Query Match   95.6%; Score 1160; DB 6; Length 225;
Best Local Similarity 96.0%; Pred. No.3.7e-105;
Matches 216; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
         |||:||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db    1  MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60
Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120
Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225
```

RESULT 4
US-10-209-208-24

;
; Sequence 24, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert

;
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 24
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "T1" with I125R mutation
US-10-209-208-24

Query Match   95.0%; Score 1153; DB 6; Length 225;
Best Local Similarity 95.6%; Pred. No. 1.8e-104;
Matches 215; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

```
Qy    1  MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFPAWDI 60
         |||:||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db    1  MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60
Qy   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGSFIY 120
Qy  121  KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  KVKFRGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
Qy  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
         |||||||||||||||||||||||||||||||||||||||||||||
Db  181  YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAEGRHHLFL 225
```

RESULT 5
US-10-209-208-6

;
; Sequence 6, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 6
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "dimer2"
US-10-209-208-6

Query Match   92.3%; Score 1121; DB 6; Length 226;

```
Best Local Similarity  92.8%;  Pred. No. 2.2e-101;
Matches 207;  Conservative 7;  Mismatches 9;  Indels 0;  Gaps 0;

Qy   3 SSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILS 62
          ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 SSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS 63

Qy  63 PQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKV 122
          |||||||| ||||||||||||||||||||||||||||||||||||||||||| ||
Db  64 PQFQYGSKAVVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGTLIYKV 123

Qy 123 KFPIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYM 182
          ||:|||||||||||||||||||||||||||:||||||:|:|||||||||||||||:||
Db 124 KFRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYM 183

Qy 183 AKKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL 225
          |||||||||||||:|||||||||||||||:||||||:::|||||
Db 184 AKKPVQLPGYYYVDTKLDITSHNEDYTIIVEQYERSEGRHHLFL 226
```

RESULT 6
US-10-209-208-8
; Sequence 8, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "mRFP1"
US-10-209-208-8

```
Query Match  84.1%;  Score 1021;  DB 6;  Length 225;
Best Local Similarity  86.9%;  Pred. No. 1e-91;
Matches 192;  Conservative 5;  Mismatches 24;  Indels 0;  Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         | |:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  61 LSPQFQYGSKAVVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||: |||||||||||||||||||||||||||| |||||: :|||||||||||:|
Db 121 KVKLRGTNFPPDGPVMQKKTMGWEASTERLYPRDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         |||||||||||   ||:||||||||||||||:|||||:|||
Db 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIIVEQYERAEGRH 221
```

RESULT 7
US-10-209-208-79
; Sequence 79, Application US/10209208

; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DsRed polypeptide variant "mRFP1.1"
US-10-209-208-79

```
Query Match  83.4%;  Score 1012;  DB 6;  Length 225;
Best Local Similarity  86.0%;  Pred. No. 7.4e-91;
Matches 190;  Conservative 6;  Mismatches 25;  Indels 0;  Gaps 0;

Qy   1 MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDI 60
         | |:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI 60

Qy  61 LSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIY 120
         |||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  61 LSPQFMYGSKAVVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY 120

Qy 121 KVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSI 180
         |||: |||||||||||||||||||||:|||:|:|||| |||||: :|||||||||:|
Db 121 KVKLRGTNFPPDGPVMQKKTMGWEASSERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTT 180

Qy 181 YMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRH 221
         |||||||||||   ||:||||||||||||||:|||||:|||
Db 181 YMAKKPVQLPGAYKTDIKLDITSHNEDYTIIVEQYERAEGRH 221
```

Search completed: January 11, 2006, 02:12:24
Job time : 10 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 12, 2006, 12:23:38

Title: US-10-006-922A-12
Perfect Score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..........EDYTIVEQYERTBGRHHLFL 225

Scoring table: BLOSUM62
             Xgapop 10.0 , Xgapext 0.5
             Ygapop 10.0 , Ygapext 0.5
             Fgapop 6.0 , Fgapext 7.0
             Delop   6.0 , Delext   7.0

Searched: 5883141 seqs, 2842172565 3 residues
             Search time 3690 Seconds
             3466.064 Million cell updates/sec

Total number of hits satisfying chosen parameters: 85

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 69%
                  Maximum Match 100%
                  Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat_10012006_162504_19114/app_query.fasta_1.391
-DB=GenEmbl -QPMT=fastap -SUFIX=p2nminscr.rge -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=500 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=69 -ALIGN=500
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_1.7415 @runat_10012006_162504_19114 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
             1: gb_ba:*
             2: gb_in:*
             3: gb_en:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pr:*
             9: gb_ro:*
            10: gb_sts:*
            11: gb_sy:*
            12: gb_un:*
            13: gb_vi:*
            14: gb_htg:*
            15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1214	100.0	678	6	AR183915	AR183915 Sequence
2	1214	100.0	678	6	AX172854	AX172854 Sequence
3	1214	100.0	678	6	AX207715	AX207715 Sequence
4	1214	100.0	678	6	AX233581	AX233581 Sequence
5	1214	100.0	678	6	AX370404	AX370404 Sequence
6	1214	100.0	678	6	AX824725	AX824725 Sequence
7	1214	100.0	859	2	AF168419	AF168419 Discosoma
8	1214	100.0	859	6	AR636082	AR636082 Sequence
9	1214	100.0	859	6	AR670157	AR670157 Sequence
10	1214	100.0	681	6	AX463698	AX463698 Sequence
11	1210	99.7	713	6	AX233584	AX233584 Sequence
12	1210	99.7	723	6	AX233627	AX233627 Sequence
13	1210	99.7	723	2	AR527331	AR527331 Sequence
14	1210	99.7	2721	6	CQ882115	CQ882115 Sequence
15	1210	99.7	2772	6	CQ882117	CQ882117 Sequence
16	1210	99.7	4692	6	AX463702	AX463702 Sequence
17	1210	99.7	711	2	AX823860	AX823860 Discosoma
18	1210	99.7	678	6	AY569779	AY569779 Cloning v
19	1210	99.4	9320	11	AX663075	AX663075 Sequence
20	1210	99.4	678	6	AY569780	AY569780 Cloning v
21	1207	98.5	666	6	AX370406	AX370406 Sequence
22	1196	98.5	666	6	AX348043	AX348043 Sequence
23	1196	98.5	711	2	AX353910	AX353910 Sequence
24	1196	98.3	678	2	AY679107	AY679107 Discosoma
25	1193	98.1	678	11	AF545828	AF545828 Discosoma
26	1191	98.1	678	6	DQ005468	DQ005468 Cloning v
27	1191	98.1	921	2	AX370408	AX370408 Sequence
28	1191	98.1	675	6	AX679106	AX679106 Discosoma
29	1186	97.7	675	6	AX824731	AX824731 Sequence
30	1186	97.7	1050	6	AX666133	AX666133 Sequence
31	1186	97.7	4488	6	CQ849509	CQ849509 Sequence
32	1186	97.7	4546	6	AY818375	AY818375 Cloning v
33	1186	97.7	4555	11	DQ005468	DQ005468 Cloning v
34	1186	97.7	4570	11	AY818373	AY818373 Cloning v
35	1186	97.7	4829	11	AY640628	AY640628 SiRNA vec
36	1186	97.7	4906	11	AY640625	AY640625 SiRNA vec
37	1186	97.7	5311	11	AJ851284	AJ851284 Cloning v
38	1186	97.7	6423	11	AY613997	AY613997 Cloning v
39	1186	97.7	6990	11	AY640634	AY640634 SiRNA vec
40	1186	97.7	7058	11	AY640633	AY640633 SiRNA vec
41	1186	97.7	7147	11	AY640630	AY640630 SiRNA vec
42	1186	97.7	7616	6	CQ849511	CQ849511 Sequence
43	1186	97.7	9468	11	AY916793	AY916793 Dual fluo
44	1186	97.7	10140	6	CS018252	CS018252 Sequence
45	1186	97.7	10522	6	CS018253	CS018253 Sequence
46	1186	97.7	10786	6	CS018254	CS018254 Sequence
47	1186	97.7	11251	6	CS018256	CS018256 Sequence
48	1186	97.7	11570	6	CS018251	CS018251 Sequence
49	1186	97.7	11867	6	CS018250	CS018250 Sequence
50	1186	97.7	11920	6	CS018255	CS018255 Sequence
51	1186	97.7	14720	6	CS018257	CS018257 Sequence
52	1186	97.7	16157	6	CQ981073	CQ981073 Sequence
53	1184	97.5	16157	6	CQ981074	CQ981074 Sequence
54	1184	97.5	898	2	AX686888	AX686888 Sequence
55	1170	96.4	3441	11	AY916793	AY916793 Dual fluo
56	1170	96.4	5311	11	AJ851285	AJ851285 Cloning v
57	1170	96.4	13079	11	AJ851286	AJ851286 Cloning v
58	1165	96.0	5311	11	AJ851289	AJ851289 Cloning v
59	1165	96.0	10141	11	AY342347	AY342347 Red H-Pel
60	1165	96.0	10276	11	AY342348	AY342348 Red H-Sti
61	1165	96.0	10481	11	AY490568	AY490568 UAS-Red S
62	1165	96.0	13079	11	AJ851290	AJ851290 Cloning v
63	1121	92.3	681	11	AF506025	AF506025 Synthetic
64	1121	92.3	1395	11	AF506026	AF506026 Synthetic
65	1119	92.2	678	11	AX824729	AX824729 Sequence
66	1095	90.2	705	11	AY678268	AY678268 Synthetic
67	1095	90.2	1431	11	AY678269	AY678269 Synthetic
68	1085.5	89.4	876	2	AF272711	AF272711 Discosoma
69	1085.5	89.4	876	2	AX686894	AX686894 Sequence
70	1021	84.1	678	11	AF506027	AF506027 Synthetic
71	1021	84.1	696	11	AB166761	AB166761 Synthetic
72	1021	84.1	4543	11	DQ005472	DQ005472 Cloning v
73	1021	84.1	4555	11	DQ005474	DQ005474 Cloning v
74	1021	84.1	4558	11	DQ005475	DQ005475 Cloning v
75	1021	84.1	5311	11	AJ851287	AJ851287 Cloning v
76	1021	84.1	13079	11	AJ851291	AJ851291 Cloning v

	Score		Length		Accession	Type
77	1007	82.9	678	11	AY678271	Synthetic
78	1002	82.5	678	11	AY678270	Synthetic
79	994	81.9	702	11	AY679163	Synthetic
80	989	81.5	711	11	AY678265	Synthetic
81	987	81.3	711	11	AY678264	Synthetic
80	985	81.1	681	2	AY785536	Discosoma
82	985	81.1	711	11	AY678266	Discosoma
83	973	80.1	711	11	AY678266	Synthetic
84	972	80.1	681	2	AY785537	Discosoma
85	960	79.1	711	11	AY678267	Synthetic

RESULT 1

```
LOCUS       AR183915                 678 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 6 from patent US 6342379.
ACCESSION   AR183915
VERSION     AR183915.1  GI:2022884
KEYWORDS    
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Tsien,R.Y. and Gonzalez,J.E. III.
  TITLE     Detection of transmembrane potentials by optical methods
  JOURNAL   Patent: US 6342379-A 6 29-JAN-2002;
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

ALIGNMENTS

```
Alignment Scores:
Pred. No.:
Score:                    1214.00
Percent Similarity:       100.00%
Best Local Similarity:    100.00%
Query Match:              6

                          Length:        678
                          Matches:       225
                          Conservative:  0
                          Mismatches:    0
                          Indels:        0
                          Gaps:          0

US-10-006-922A-12 (1-225) x AR183915 (1-678)

Qy    1  MetArgSerSerLysAsnValIleLysGluMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTGAGGTTCAAGGTTCGCATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAGGGCGAAGGCGAGGGGAGGCCATACGAAGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCCTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCAGATATCCCA  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACAGAGCGTCTGTATCCTCGTGATGGCGTTGTCTTGAAAGGAGAG  480

Qy  161  IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTTGAATTCAAAAGTATT  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCCGTTCAGCTACCAGGTTACTACTATGTTGACTCCAAACTTGAT  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGAAGC  660

Qy  221  HisHisLeuPheLeu  225
         |||||||||||||||
Db  661  CACCATCTGTTCCTT  675
```

RESULT 2

```
LOCUS       AX172854                 678 bp    DNA     linear   PAT 03-JUL-2001
DEFINITION  Sequence 6 from Patent WO0142211.
ACCESSION   AX172854
VERSION     AX172854.1  GI:14597903
KEYWORDS    
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Tsien,R.Y. and Gonzalez,J.E.
  TITLE     Detection of transmembrane potentials by optical methods
  JOURNAL   Patent: WO 0142211-A 6 14-JUN-2001;
            The Regents of the University of California (US)
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                    1214.00
Percent Similarity:       100.00%
Best Local Similarity:    100.00%
Query Match:              6

                          Length:        678
                          Matches:       225
                          Conservative:  0
                          Mismatches:    0
                          Indels:        0
                          Gaps:          0

US-10-006-922A-12 (1-225) x AX172854 (1-678)

Qy    1  MetArgSerSerLysAsnValIleLysGluMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTGAGGTTCAAGGTTCGCATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAGGGCGAAGGCGAGGGGAGGCCATACGAAGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCCTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCAGATATCCCA  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  360
```

```
RESULT 3

LOCUS       AX207715                 678 bp    DNA     linear   PAT 31-AUG-2001
DEFINITION  Sequence 13 from Patent WO0157242.
ACCESSION   AX207715
VERSION     AX207715.1  GI:15422399
KEYWORDS
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Stack,J.H., Whitney,M., Cubitt,A.B. and Pollok,B.A.
  TITLE     Methods of protein destabilization and uses thereof
  JOURNAL   Patent: WO 0157242-A 13 09-AUG-2001;
            Aurora Biosciences Corporation (US)
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"

US-10-006-922A-12 (1-225) x AX207715 (1-678)

Alignment Scores:
  Pred. No.:
  Score:                     1214.00         Length:           678
  Percent Similarity:        100.00%         Matches:          225
  Best Local Similarity:     100.00%         Conservative:       0
  Query Match:               100.00%         Mismatches:         0
  DB:                                        Indels:             0
                                    6        Gaps:               0

ORIGIN

Qy    1  MetArgSerLysAsnValIleLysGlyLeuPheMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTAAAGGTTCGCATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGGAGACCATACGAAGGC 120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 4

LOCUS       AX233581                 678 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 5 from Patent WO0162919.
ACCESSION   AX233581
VERSION     AX233581.1  GI:15593305
KEYWORDS
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Nelson,D., Zamaira,E. and Tsien,R.
  TITLE     Modified fluorescent proteins
  JOURNAL   Patent: WO 0162919-A 5 30-AUG-2001;
            Aurora Biosciences Corporation (US)
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"
                     /note="'red'"

US-10-006-922A-12 (1-225) x AX233581 (1-678)

Alignment Scores:
  Pred. No.:
  Score:                     1214.00         Length:           678
  Percent Similarity:        100.00%         Matches:          225
  Best Local Similarity:     100.00%         Conservative:       0
  Query Match:               100.00%         Mismatches:         0
  DB:                                        Indels:             0
                                    6        Gaps:               0

ORIGIN

Qy    1  MetArgSerLysAsnValIleLysGlyLeuPheMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTAAAGGTTCGCATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
```

```
RESULT 5
AX370404
LOCUS       AX370404          678 bp    DNA    linear   PAT 16-FEB-2002
DEFINITION  Sequence 1 from Patent WO0196373.
ACCESSION   AX370404
VERSION     AX370404.1  GI:18857490
KEYWORDS
SOURCE
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
            synthetic construct
            synthetic construct
REFERENCE   1
  AUTHORS   Pradkov,A.F. and Terskikh,A.
  TITLE     Fluorescent timer proteins and methods for their use
  JOURNAL   Patent: WO 0196373-A 1 20-DEC-2001;
            Clontech Laboratories Inc. (US)
FEATURES            Location/Qualifiers
     source         1..678
                    /organism="synthetic construct"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:32630"
                    /note="variant of sequence from Discosoma sp."
ORIGIN
```

```
Alignment Scores:

Pred. No.:              1.4e-114
Score:                  1214.00
Percent Similarity:     100.00%
Best Local Similarity:  100.00%
Query Match:            100.00%
DB:                     6

Length:     678
Matches:    225
Conservative: 0
Mismatches: 0
Indels:     0
Gaps:       0
```

```
Qy   41  HisAsnThrValLysLeuSerValThrLysGlyGlyGlyLeuProPheAlaThrTrpAspIle  60
Db   61  ACGGTCAATGGCACGAGTTGAAATAGAAGAGGAAGGCCATACCGGAAGGC             120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro     80
Db  121  CACAATACCCGTAAAGCTTAAGGTAACCAAGGGGGGGACCTTTGCCATTTGGGTATATT     180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheProGluGlyValMetAsnPheGlu        100
Db  181  TTGTCACCACAATTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA     240
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr      120
Db  241  GACTATAAAAAGCTTCATTTCCTGAAGGAGGATTTAAAATGGGGAAAGTTTTGAA          300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr      120
Db  301  GACGGCGTGGGCGTTGCTGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC     360
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr      140
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA     420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu     160
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTACCCTCGTGATGGCGTGTTGAAGGGAGAG     480
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  161  IleHisGlyAlaLeuLysLeuGlyAspGlyGlyHisTyrLeuValGluPheLysSerIle     180
Db  481  ATTCATGGAGCCCTGAAGCTGGGAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT     540
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp     200
Db  541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTATTATGTTGACTCCAAACTGGAT     600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg     220
Db  601  ATAACAAGCCACAACGAAGACTATACAATTGTGGAGCAGTATGAAAGAACCGAGGGACGC     660
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  221  HisHisLeuPheLeu                                                  225
Db  661  CACCACCTGTTCCTT                                                 675
         |||||||||||||||
```

```
US-10-006-922a-12 (1-225) x AX370404 (1-678)
```

```
RESULT 6
AX824725
LOCUS       AX824725          678 bp    DNA    linear   PAT 11-DEC-2003
DEFINITION  Sequence 7 from Patent WO02068459.
ACCESSION   AX824725
VERSION     AX824725.1  GI:39750591
KEYWORDS
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS
  TITLE     Non aggregating fluorescent proteins and methods for using the same
  JOURNAL   Patent: WO 02068459-A 7 06-SEP-2002;
FEATURES            Location/Qualifiers
     source         1..678
                    /organism="Discosoma sp."
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:86600"
ORIGIN
```

```
Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly     20
Db    1  ATGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGGAGGC     60
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly     40
Db   61  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC     120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   41  AspTyrLysLysLeuSerPheProGluGlyPheProGluGlyValMetAsnPheGlu        60
Db  121  CACAACACCCGTGAAGCTGAAGGTGACCAAGGGGGGGACCTTTGCCATTTGGGACATC       180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro     80
Db  181  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC     240
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu     100
Db  241  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG     300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr     120
Db  301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC     360
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr     140
Db  361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC     420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu     160
Db  421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG     480
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  161  IleHisGlyAlaLeuLysLeuGlyAspGlyGlyHisTyrLeuValGluPheLysSerIle     180
Db  481  ATCCACGGCGCCCTGAAGCTGGGCGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC     540
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp     200
Db  541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC     600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg     220
Db  601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC     660
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  221  HisHisLeuPheLeu                                                  225
Db  661  CACCACCTGTTCCTG                                                 675
         |||||||||||||||
```

Alignment Scores:
Pred. No.:   1
Score: 1.4e-114
                        1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 6

Length: 678
Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AX824725 (1-678)

```
Qy   1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCGCTCCAAGAACGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy  81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 221 HisHisHisLeuPheLeu 225
        ||||||||||||||||||
Db 661 CACCACCTGTTCCTG 675
```

RESULT 7

LOCUS   AF168419   859 bp   mRNA   linear   INV 27-JUL-2001
DEFINITION Discosoma sp. fluorescent protein FP583 mRNA, complete cds.
ACCESSION   AF168419
VERSION   AF168419.2   GI:7105733
KEYWORDS   .
SOURCE   Discosoma sp.
  ORGANISM Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1 (bases 1 to 859)
  AUTHORS   Matz,M.V., Fradkov,A.P., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
            Markelov,M.L. and Lukyanov,S.A.
  TITLE   Fluorescent proteins from nonbioluminescent Anthozoa species

JOURNAL Nat. Biotechnol. 17 (10), 969-973 (1999)
REFERENCE 2 (bases 1 to 859)
  PUBMED   10504696
  AUTHORS Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
           Markelov,M.L. and Lukyanov,S.A.
  TITLE   Direct Submission
  JOURNAL Submitted (14-JUL-1999) Institute of Bioorganic Chemistry RAS,
           Miklukho-Maklaya 16/10, Moscow 117871, Russia
REFERENCE 3 (bases 1 to 859)
  AUTHORS Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
           Markelov,M.L. and Lukyanov,S.A.
  TITLE   Direct Submission
  JOURNAL Submitted (25-FEB-2000) Institute of Bioorganic Chemistry RAS,
           Miklukho-Maklaya 16/10, Moscow 117871, Russia
REMARK   Sequence update by submitter
COMMENT   On Feb 25, 2000 this sequence version replaced gi:6090866.
FEATURES   Location/Qualifiers
  source   1..859
                  /organism="Discosoma sp."
                  /mol_type="mRNA"
                  /db_xref="taxon:86600"
  CDS   54..731
                  /note="GFP-like protein; orange-red"
                  /codon_start=1
                  /product="fluorescent protein FP583"
                  /protein_id="AAF03369.1"
                  /db_xref="GI:6090867"
                  /translation="MRSSKNVIKEFMRFKVRMEGTVNGHEFEISEGEGRPYEGHNTV
                  KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                  GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                  IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
                  GRHHLFL"
ORIGIN

Alignment Scores:
Pred. No.:   1
Score: 1.87e-114
                        1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 2

Length: 859
Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AF168419 (1-859)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAAAGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAACGGAAGC 173

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGAGGACCTTTGCCATTTGCCTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAACGGGTAATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACGGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
```

Left column

```
RESULT 10

LOCUS         AX463698              859 bp    DNA        linear   PAT 15-JUL-2002
DEFINITION    Sequence 12 from Patent WO0248338.
ACCESSION     AX463698
VERSION       AX463698.1  GI:21886457
KEYWORDS
SOURCE        Discosoma sp.
  ORGANISM    Discosoma sp.
              Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
              Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE     1
  AUTHORS     Lichtenberg-Frat,H.
  TITLE       Yeast strain for testing the geno- and cytotoxicity of complex
              environmental contamination
  JOURNAL     Patent: WO 0248338-A 12 20-JUN-2002;
              Lichtenberg-Frat, Hella (DE)
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"
     CDS             54..731
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD42147.1"
                     /db_xref="GI:21886458"
                     /translation="MRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
```

Alignment Scores:
```
Pred. No.:              1
Score:                  1.87e-114           Length:              859
Percent Similarity:     1214.00             Matches:             225
Best Local Similarity:  54..731             Conservative:        0
Query Match:            100.00%             Mismatches:          0
DB:                     100.00%             Indels:              0
                        100.00%             Gaps:                0
                        6
```

Right column

US-10-006-922A-12 (1-225) x AX463698 (1-859)

```
RESULT 11

LOCUS         AX233584              681 bp    DNA        linear   PAT 11-SEP-2001
DEFINITION    Sequence 8 from Patent WO0162919.
ACCESSION     AX233584
VERSION       AX233584.1  GI:15593307
KEYWORDS
SOURCE        synthetic construct
  ORGANISM    synthetic construct
              other sequences; artificial sequences.
REFERENCE     1
  AUTHORS     Nelson,D., Zamaira,E. and Tsien,R.
  TITLE       Modified fluorescent proteins
  JOURNAL     Patent: WO 0162919-A 8 30-AUG-2001;
              Aurora Biosciences Corporation (US)
FEATURES             Location/Qualifiers
     source          1..681
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     CDS             1..681
                     /note="Mutant Anthozoan red fluorescent protein"
                     /note="unnamed protein product"
```

```
ORIGIN
        /codon_start=1
        /transl_table=11
        /protein_id="CAC69734.1"
        /db_xref="GI:15593308"
        /translation="MVRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNT
        VKLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFED
        GGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKG
        EIHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERT
        EGRHHLFL"
```

US-10-006-922A-12 (1-225) x AX233584 (1-681)

```
Alignment Scores:
Pred. No.:
Score:                    3.61e-114       Length:         681
Percent Similarity:       1210.00         Matches:        224
Best Local Similarity:    100.00%         Conservative:   1
                          99.56%          Mismatches:     0
Query Match:              99.67%          Indels:         0
DB:                                       Gaps:           0
```

```
Qy    1 MetArgSerSerLysSerAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4 GTGAGGAGCAGCAAGAAACGTGATCAAGGAGTTCATGAGGTTCAAGGTGCGCATGGAGGGC   63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly    40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle    60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro    80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184 CTGAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   243

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  244 GACTACAAGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   303

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  304 GACGGCGGGGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCTGCTTCATCTAC   363

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGGCCCGTGATGCAGAAGAAGACC   423

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGGGTGCTGAAGGGCGAG   483

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  484 ATCCACAAGGCCCTGAAGCTGAAGGACGGGGGCCACTACCTGGTGGAGTTCAAGTCCATC   543

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC   603

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  604 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGAGGACCGAGGGGCAGG   663

Qy  221 HisHisLeuPheLeu   225
        |||||||||||||||
Db  664 CACCACCTGTTCCTG   678
```

```
LOCUS       AX233627                 713 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 51 from Patent WO0162919.
ACCESSION   AX233627
VERSION     AX233627.1  GI:15593330
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Nelson,D., Zamaira,E. and Tsien,R.
  TITLE     Modified fluorescent proteins
  JOURNAL   Patent: WO 0162919-A 51 30-AUG-2001;
            Aurora Biosciences Corporation (US)
FEATURES             Location/Qualifiers
     source          1..713
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Mutant Anthozoan red fluorescent protein"
ORIGIN
```

US-10-006-922A-12 (1-225) x AX233627 (1-713)

RESULT 12
AX233627

```
Alignment Scores:
Pred. No.:
Score:                    3.81e-114       Length:         713
Percent Similarity:       1210.00         Matches:        224
Best Local Similarity:    100.00%         Conservative:   1
                          99.56%          Mismatches:     0
Query Match:              99.67%          Indels:         0
DB:                                       Gaps:           0
```

```
Qy    1 MetArgSerSerLysSerAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
        ::::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 GTGAGGAGCAGCAAGAAACGTGATCAAGGAGTTCATGAGGTTCAAGGTGCGCATGGAGGGC   81

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly    40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   141

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle    60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   201

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro    80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202 CTGAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   261

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262 GACTACAAGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   321

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322 GACGGCGGGGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCTGCTTCATCTAC   381

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGGCCCGTGATGCAGAAGAAGACC   441

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  442 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGGGTGCTGAAGGGCGAG   501

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  502 ATCCACAAGGCCCTGAAGCTGAAGGACGGGGGCCACTACCTGGTGGAGTTCAAGTCCATC   561

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  562 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC   621

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGAGGACCGAGGGGCAGG   681
```

```
Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db    682 CACCACCTGTTCCTG 696
```

RESULT 13

```
Db         AR527331
LOCUS      AR527331              723 bp    DNA        linear   PAT 08-OCT-2004
DEFINITION Sequence 1 from patent US 6723537.
ACCESSION  AR527331
VERSION    AR527331.1  GI:53914309
KEYWORDS   .
SOURCE     Unknown.
  ORGANISM Unknown.
           Unclassified.
REFERENCE  1  (bases 1 to 723)
  AUTHORS  Peelle,B.
  TITLE    Directed evolution of protein in mammalian cells
  JOURNAL  Patent: US 672537-A 1 20-APR-2004;
           Rigel Pharmaceuticals, Incorporated and Becton, Dickinson and
           Company; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1..723
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

US-10-006-922A-12 (1-225) x AR527331 (1-723)

```
Alignment Scores:
Pred. No.:
Score:                1210.00
Percent Similarity:   100.00%
Best Local Similarity: 99.56%
Query Match:          6
DB:
```

Length: 723
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

Score: 3.88e-114

```
Qy    1   MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly 20
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4   GTGGCGCTTCCTCCAAGGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy    81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 GACTACAAGAGACTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy    101 AspGlyGlyValValThrGlnAspSerSerLeuGlnAspGlyGlyTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GACGGCGGCGTGGTGACCCAGGACCTCCTCCCTGCAGGACGGCGGCTTCATCTAC 363

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTAATCCAGACGGCCCGGTCATGCAGAAGAAGACC 423

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTCCTGAAGGGCGAG 483

Qy    161 IleHisGlyAlaLeuLysLysAspGlyGlyHisHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 ATCCACGGCGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGTATC 543

Qy    181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         ...
```

```
Db    544 TACATGGCCAAGGAAGCCCGTGCAGCTGCCCGGCTACTACTAGGTGGACTCCAAGCTGGAC 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db    664 CACCACCTGTTCCTG 678
```

RESULT 14

```
Db         CQ882115
LOCUS      CQ882115             2721 bp    DNA        linear   PAT 11-OCT-2004
DEFINITION Sequence 1 from Patent WO2004083445.
ACCESSION  CQ882115
VERSION    CQ882115.1  GI:54034825
KEYWORDS   .
SOURCE     synthetic construct
  ORGANISM synthetic construct
           other sequences; artificial sequences.
REFERENCE  1
  AUTHORS  Chavancy,G., Couble,P., Durand,B., Grenier,A.M., Horard,B.,
           Julien,E., Mauchamp,B., Nony,P., Prudhomme,J.C. and Royer,C.
  TITLE    Nucleic acid controlling the expression of a useful polypeptide in
           the posterior silk glands of a lepidoptera and application thereof
  JOURNAL  Patent: WO 2004083445-A1 30-SEP-2004;
           Centre National De La Recherche Scientifique-CNRS (FR); UNIVERSITE
           CLAUDE BERNARD - LYON 1 (FR); Institut National de la Recherche
           Agronomique (INRA) (FR)
FEATURES             Location/Qualifiers
     source          1..2721
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description de la s quence artificielle : s quence
                     de fusion"
ORIGIN
```

US-10-006-922A-12 (1-225) x CQ882115 (1-2721)

```
Alignment Scores:
Pred. No.:
Score:                1210.00
Percent Similarity:   100.00%
Best Local Similarity: 99.56%
Query Match:          6
DB1:
```

Length: 2721
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

Score: 1.95e-113

```
Qy    1    MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly 20
           :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2044 GTGGCGCTCCTCCAAGGAACCTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 2103

Qy    21   ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2104 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2163

Qy    41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2164 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2223

Qy    61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2224 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2283

Qy    81   AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2284 GACTACAAGAAAGCTGTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2343

Qy    101  AspGlyGlyValValThrGlnAspSerSerLeuGlnAspGlyGlyTyrLysCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2344 GACGGCGGCGTGGTGGTGACCCAGGACAGCTCCCTGCAGGACGGCGGCTTCGCTTCATCTAC 2403
```

```
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyValMetGlnLysGlyLysThr 140
Db 2404 AAGGTGAAGTTCATCGGCGTGAACTTCCCGACGGCCCGTAATGCAAGAAGAGACC 2463
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProAlaArgAspGlyValLeuLysGlyGlu 160
Db 2464 ATGGGCTGGGAAGCGTCCACGGAGCGCCTGTACCCCGCCCGTGATGGCGTGCTGAAGGGCGAG 2523
Qy  161 IleHisHisLysAlaLeuLeuLysLysLysAspGlyGlyHisTyrLysTyrSerIle 180
Db 2524 ATCCACAAGGCCCTGAAGCTGAAGAAGAAGGATGGTGGCCACTACAAGTACTCCATC 2583
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 2584 TACAATGGCCAAGAAGACCCGTGCAGCTGCCGGGCTACTACTACGTGGAC 2643
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 2644 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2703
Qy  221 HisHisLeuPheLeu 225
Db 2704 CACCACCTGTTCCTG 2718
```

RESULT 15
CQ882117

```
LOCUS       CQ882117                2772 bp    DNA     linear   PAT 11-OCT-2004
DEFINITION  Sequence 3 from Patent WO2004083445.
ACCESSION   CQ882117
VERSION     CQ882117.1  GI:54034827
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (residues 1 to 2772)
  AUTHORS   Chavancy,G., Couble,P., Durand,B., Grenier,A.M., Horard,B.,
            Julien,E., Mauchamp,B., Nony,P., Prudhomme,J.C. and Royer,C.
  TITLE     Nucleic acid controlling the expression of a useful polypeptide in
            the posterior silk glands of a lepidoptera and application thereof
  JOURNAL   Patent: WO 2004083445-A3 30-SEP-2004;
            Centre National De La Recherche Scientifique-CNRS (FR); UNIVERSITE
            CLAUDE BERNARD - LYON 1 (FR); Institut National de la Recherche
            Agronomique (INRA) (FR)
FEATURES             Location/Qualifiers
     source          1..2772
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description de la s quence artificielle : s quence
                     de fusion"
ORIGIN
```

US-10-006-922A-12 (1-225) x CQ882117 (1-2772)

```
Alignment Scores:
Pred. No.:                              Length:        2772
Score:                  1210.00         Matches:        224
Percent Similarity:      100.00%        Conservative:     1
Best Local Similarity:    99.56%        Mismatches:       0
Query Match:              99.67%        Indels:           0
DB:                           6         Gaps:             0
                                        2e-113
```

US-10-006-922A-12 (1-225) x CQ882117 (1-2772)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
        :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2095 GTGCGCTTCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 2154
Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2155 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2214
Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2215 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2274
```

```
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2275 CTGTCCCCACAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCGGCCAACATCCCC 2334
Qy   81 AspGlyTyrLysLysSerPheProGluGlyGlyProPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2335 GACTACAAGAAGAGCTTCCCCGAGGGCGGCCCCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 2394
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2395 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 2454
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2455 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 2514
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2515 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2574
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2575 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 2634
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2635 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2694
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2695 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2754
Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 2755 CACCACCTGTTCCTG 2769
```

RESULT 16
AX463702

```
LOCUS       AX463702                4692 bp    DNA     linear   PAT 15-JUL-2002
DEFINITION  Sequence 16 from Patent WO0248338.
ACCESSION   AX463702
VERSION     AX463702.1  GI:21886461
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (residues 1 to 4692)
  AUTHORS   Lichtenberg-Frat,H.
  TITLE     Yeast strain for testing the geno- and cytotoxicity of complex
            environmental contamination
  JOURNAL   Patent: WO 0248338-A 16 20-JUN-2002;
            Lichtenberg-Frat, Hella (DE)
FEATURES             Location/Qualifiers
     source          1..4692
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Vektor pDsRed1-N1"
ORIGIN
```

US-10-006-922A-12 (1-225) x AX463702 (1-4692)

```
Alignment Scores:
Pred. No.:                              Length:        4692
Score:                  1210.00         Matches:        224
Percent Similarity:      100.00%        Conservative:     1
Best Local Similarity:    99.56%        Mismatches:       0
Query Match:              99.67%        Indels:           0
DB:                                     Gaps:             0
                                        3.79e-113
```

US-10-006-922A-12 (1-225) x AX463702 (1-4692)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
        :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 17

```
LOCUS        AX823860              6893 bp    DNA        linear    PAT 11-DEC-2003
DEFINITION   Sequence 8 from Patent WO03070931.
ACCESSION    AX823860
VERSION      AX823860.1  GI:39750176
KEYWORDS     .
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1
  AUTHORS    Baum,C., Will,E., Osterag,W., Klump,H. and Schiedlmeier,B.
  TITLE      Methods for conducting site-specific dna recombination
  JOURNAL    Patent: WO 0307093l-A 8 28-AUG-2003;
             Vision 7 GmbH (DE)
FEATURES             Location/Qualifiers
     source          1..6893
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Beschreibung der kuenstlichen Sequenz: provirale
                     Plasmid-DNA; retrovirales Cre-Reporterkonstrukt #SR#"
     misc_feature    1..435
                     /note="plasmid-Rueckgrat (pUC)"
     misc_feature    436..996
                     /note="MPSV 5'-LTR (Delta1-31)"
     primer_bind     997..1014
```

US-10-006-922A-12 (1-225) x AX823860 (1-6893)

Alignment Scores:
Pred. No.:
Score: 1210.00
Percent Similarity: 100.00%
Best Local Similarity: 99.56%
Query Match: 99.67%
DB: 6

Length: 6893
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

[Alignment / sequence data follows — paired query (Qy) and database (Db) nucleotide and amino-acid translation lines with position numbers from 1 to ~2262 on the left block and 682 to 1356 on the right block.]

```
RESULT 18

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||
Db 2263 CACCACCTGTTCCTG 2277

LOCUS       AY569779/c              8811 bp    DNA     circular SYN 05-APR-2004
DEFINITION  Cloning vector pTCCR, complete sequence.
ACCESSION   AY569779
VERSION     AY569779.1  GI:45861232
KEYWORDS    .
SOURCE      Cloning vector pTCCR
  ORGANISM  Cloning vector pTCCR
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 8811)
  AUTHORS   Wu,G.; Wang,Y. and Zhu,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2004) Fish Genetics, Institute of Hydrobiology,
            Chinese Academy of Science, 7# South Road of East Lake, Wuhan,
            Hubei 430072, People's Republic of China
FEATURES             Location/Qualifiers
     source          1..8811
                     /organism="Cloning vector pTCCR"
                     /mol_type="other DNA"
                     /db_xref="taxon:268733"
     misc_feature    419..452
                     /note="LoxP RE"
     CDS             complement(505..555)
                     /codon_start=1
                     /transl_table=11
                     /product="RFP"
                     /protein_id="AAS78495.1"
                     /db_xref="GI:45861233"
                     /translation="MVRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNT
                     VKLKVTKGGPLPFAWDILSPQFQYGSKYYVKHPADIPDYKKLSFPEGFKWERVMNFED
                     GGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKG
                     EIHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERT
                     EGRHHLFL"
     terminator      complement(707..1387)
                     /note="SV40 polyA"
     promoter        complement(1447..2036)
                     /note="CMV promoter"
     CDS             complement(2387..2761)
                     /codon_start=1
                     /transl_table=11
                     /product="zeocin"
                     /protein_id="AAS78496.1"
                     /db_xref="GI:45861234"
                     /translation="MAKLTSAVPVLTARDVAGAVEFWTDRLGFSRDFVEDDFAGVVRD
                     DVTLFISAVQDQVPDNTLAWVWVRGLDELYAEWSEVVSTNFRDASGPAMTEIGEQPW
                     GREFALRDPAGNCVHFVAEEQD"
     terminator      complement(2860..2992)
                     /note="SV40 polyA"
     CDS             complement(3066..4097)
                     /codon_start=1
                     /product="Cre"
                     /transl_table=11
                     /protein_id="AAS78497.1"
                     /db_xref="GI:45861235"
                     /translation="MANLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKM
                     LLSVCRSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRS
                     GLPRPSDSNAVSIVMRRIRKENVDAGERAKQALAFERTDFQVRSLMENSDRCQDIRN
                     LAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVT
                     KLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAK
                     DDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGA
                     MVRLLEDGD"
     promoter        complement(4109..4683)
     misc_feature    complement(4745..4762)
                     /note="internal ribosome entry sequence from
                     encephalomyocarditis virus"
     promoter        complement(5151..6233)
                     /note="T7 promoter"
     CDS             
                     /product="LacI"
                     /protein_id="AAS78498.1"
                     /db_xref="GI:45861236"
                     /translation="MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAE
                     LNYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAITKSRADQLGASVVVSMVERS
                     GVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSII
                     FSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIPIAEREG
                     DWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDT
                     EDSSCYIPPSTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNT
                     QTASPRALADSLMQLARQVSRLESGQ"
     misc_feature    6520..6553
                     /note="LoxP LE"
     rep_origin      6992
                     /note="Col E1"
     CDS             complement(7751..8611)
                     /codon_start=1
                     /transl_table=11
                     /product="Amp"
                     /protein_id="AAS78499.1"
                     /db_xref="GI:45861237"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IEDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
ORIGIN

US-10-006-922A-12 (1-225) x AY569779 (1-8811)

Alignment Scores:
Pred. No.:            8.18e-113          Length:        8811
Score:               1210.00            Matches:        224
Percent Similarity:   100.00%           Conservative:     1
Best Local Similarity: 99.56%           Mismatches:       0
Query Match:          99.67%            Indels:           0
DB:                      11             Gaps:             0

Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1384 GTGCGGTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1325

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1324 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1265

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1264 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1205

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysTyrTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1204 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGTACTACGTGAAGCACCCCGCCGACATCCCC 1145

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1144 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1085

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1084 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1025

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1024 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 965

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  964 ATGGGCTGGGAGGCCTCCACGGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 905

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    904  ATCCACAAGGCCCTGAAGCTGAAGACGCGGCGGCGGCCCACTACCTGGTCGGAGTTCCATC  845

Qy    181  TyrMetAlaLysGlyLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    844  TACATGGCCAAGGGCAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  785

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGlnGlnTyrGluArgThrGluGlyArg  220
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784  ATCACCTCCCACAACGAGGACTACACCATCGTGCAGCAGTACGAGCGCACCGAGGGCCGC  725

Qy    221  HisHisLeuPheLeu  225
                |||||||||||||||
Db    724  CACCACCTGTTCCTG  710
```

```
RESULT 19
LOCUS        AX663075                9320 bp    DNA     linear   PAT 24-MAR-2003
DEFINITION   Sequence 20 from Patent WO0270740.
ACCESSION    AX663075
VERSION      AX663075.1  GI:29169369
KEYWORDS     .
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1
  AUTHORS    Wiesmueller,L.
  TITLE      Test system for determining gene toxicities
  JOURNAL    Patent: WO 0270740-A 20 12-SEP-2002;
             Wiesmueller, Lisa (DE)
FEATURES             Location/Qualifiers
     source          1..9320
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     misc_feature    1..1592
                     /note="Retroviraler Vektor p5NM"
     CDS             1617..2216
                     /note="unnamed protein product; Puromycin-Resistenzgen aus
                     pRetroon (Clontech, Palo Alto, CA, USA)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD80174.1"
                     /db_xref="GI:29169370"
                     /translation="MTEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHTVDPDRHIE
                     RVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPESVEAGAVFAEIGPRMAELSGSR
                     LAAQQOMEGLLAPHRPKEPAWFLATVGVSPDHQGKGLGSAVVLPGVEAAERAGVPAFL
                     ETSAPRNLPFYERLGFTVTADVECPKDRATWCMTRKPGA"
     promoter        2267..2848
                     /note="CMV-Promotor aus pEGFP-N1 (Clontech, Palo Alto, CA,
                     USA)"
     CDS             2906..>3349
                     /note="unnamed protein product; N'-EGFP, abgeleitet von
                     EGFP aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD80175.1"
                     /db_xref="GI:29169371"
                     /translation="MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT
                     LKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFK
                     DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNR"
     promoter        3374..3992
                     /note="CMV-Promotor aus pDsRed1-N1 (Clontech, Palo Alto,
                     CA, USA)"
     misc_feature    3411..3992
                     /note="Retroviraler Vektor p5NM"
     CDS             4038..4718
                     /note="unnamed protein product; Red aus pDsRed1-N1
                     (Clontech, Palo Alto, CA, USA)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD80176.1"
                     /db_xref="GI:29169372"
```

```
                     /translation="MVRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNT
                     VKLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFED
                     GGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLIYPRDGVLKG
                     EIHKALKLDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERT
                     EGRHHLFL"
     misc_feature    4766..5509
                     /note="EGFP-EJ, abgeleitet von EGFP aus pEGFP-N1
                     (Clontech, Palo Alto, CA, USA)"
     misc_feature    5527..9320
                     /note="Retroviraler Vektor p5NM"
ORIGIN
```

```
US-10-006-922A-12 (1-225) x AX663075 (1-9320)

  Alignment Scores:
  Pred. No.:
  Score:                  1210.00    Length:           9320
  Pred. No.:              8.76e-113  Matches:           224
  Percent Similarity:     100.00%    Conservative:        1
  Percent Similarity:     99.56%     Mismatches:          0
  Best Local Similarity:  99.67%     Indels:              0
  Query Match:            6          Gaps:                0
  DB:
```

```
Qy     1  MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
               :::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4041  GTGCGCTTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  4100

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4101  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC  4160

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4161  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  4220

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4221  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  4280

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4281  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG  4340

Qy   101  AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4341  GACGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  4400

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4401  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC  4460

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4461  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  4520

Qy   161  IleHisLysAlaLeuLysLeuAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4521  ATCCACAAGGCCCTGAAGCTGGACGGCGGCCACTACCTGGTCGAGTTCAAGTCCATC  4580

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4581  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  4640

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGlnGlnTyrGluArgThrGluGlyArg  220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4641  ATCACCTCCCACAACGAGGACTACACCATCGTGCAGCAGTACGAGCGCACCGAGGGCCGC  4700

Qy   221  HisHisLeuPheLeu  225
               |||||||||||||||
Db  4701  CACCACCTGTTCCTG  4715
```

```
RESULT 20
LOCUS        AY569780               12404 bp    DNA     circular SYN 05-APR-2004
             AY569780/c
```

```
DEFINITION  Cloning vector pTCCR-Auto, complete sequence.
ACCESSION   AY569780
VERSION     AY569780.1  GI:45861238
KEYWORDS    .
SOURCE      Cloning vector pTCCR-Auto
  ORGANISM  Cloning vector pTCCR-Auto
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 12404)
  AUTHORS   Wu,G., Wang,Y. and Zhu,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2004) Fish Genetics, Institute of Hydrobiology,
            Chinese Academy of Science, 7# South Road of East Lake, Wuhan,
            Hubei 430072, People's Republic of China
FEATURES             Location/Qualifiers
     source          1..12404
                     /organism="Cloning vector pTCCR-Auto"
                     /mol_type="other DNA"
                     /db_xref="taxon:268734"
     misc_feature    419..452
                     /note="LoxP RE"
     terminator      complement(505..555)
                     /note="SV40 polyA"
     CDS             complement(707..1387)
                     /codon_start=1
                     /transl_table=11
                     /product="RFP"
                     /protein_id="AAS78500.1"
                     /db_xref="GI:45861239"
                     /translation="MVRSSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNT
                     VKLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFED
                     GGVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTRLYPRDGVLKG
                     EIHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERI
                     EGRHHLFLI"
     CDS             complement(1447..2036)
                     /codon_start=1
                     /transl_table=11
                     /note="CMV promoter"
     promoter        complement(2387..2761)
                     /codon_start=1
                     /transl_table=11
                     /product="zeocin"
                     /protein_id="AAS78501.1"
                     /db_xref="GI:45861240"
                     /translation="MAKLTSAVPVLTARDVAGAVEFWTDRLGFSRDFVEDDFAGVVRD
                     DVTLFISAVQDQVVPDNTLAWVWVRGLDELYAEWSEVVSTNFRDASGPAMTEIGEQPW
                     GREFALRDPAGNCVHFVAEEQD"
     terminator      complement(2860..2990)
                     /note="T7 terminator"
     CDS             complement(3066..4097)
                     /codon_start=1
                     /transl_table=11
                     /product="Cre"
                     /protein_id="AAS78502.1"
                     /translation="MANLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKM
                     LLSVCRSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRS
                     GLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRN
                     LAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVT
                     KLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRIIYGAK
                     DDSGQRYLAWSGHSARVCAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGA
                     MVRLLEDGD"
     misc_feature    complement(4109..4683)
                     /note="internal ribosome entry sequence from
                     encephalomyocarditis virus"
     promoter        complement(4745..4762)
                     /note="T7 promoter"
     terminator      complement(4837..4967)
                     /note="T7 terminator"
     CDS             complement(5009..7660)
                     /codon_start=1
                     /transl_table=11
                     /product="T7 RNA polymerase"
                     /protein_id="AAS78503.1"
                     /db_xref="GI:45861242"
                     /translation="MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESY
                     EMGEARFRKMFERQLKAGEVADNAAAKPLITTLLPKMIARINDWFEEVKARGIKPTA
                     FQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAVASAIGRAIEDEARFGRIRDLEAK
                     HFKKNVEEQLNKRVGHVYKKAFMQVVEADMLSKGLLGGEAWSSWHKEDSIHVGVRCIE
                     MLIESTGMVSLHRQNAGVVGQDSETIELAPEYAEAIATRAGALAGISPMFQPCVVPPK
                     PWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYKAINIAQNTAWKINKKV
                     LAVANVITKWKHCPVEDIPAIEREELPMKPEDIDMNPEALTAWKRAAAAVYRKDARK
                     SRISLEFMLEQANKFANHKAIWFPYNMDWRGRVYAVSMFNPQGNDMTKGLLTLAKGK
                     PIGKEGYYWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSP
                     FCFLAFCFEYAGVQHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSET
                     VQDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYG
                     VTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKLIW
                     ESVSVTVAAVEANMLKSAKLLAAEVKDKKTGEILRKRCAVHWVTPDGFPVWQEYK
                     KPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWA
                     HEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFYDQFADQLHES
                     QLDKMPALPAKGNLNLRDILESDFAFA"
     misc_signal     complement(7670..7690)
                     /note="SV40 T-antigen NLS"
     misc_feature    complement(7702..8276)
                     /note="internal ribosome entry sequence from
                     encephalomyocarditis virus"
     promoter        complement(8338..8355)
                     /note="T7 promoter"
     CDS             complement(8744..9826)
                     /codon_start=1
                     /transl_table=11
                     /product="LacI"
                     /protein_id="AAS78504.1"
                     /db_xref="GI:45861243"
                     /translation="MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAE
                     LNYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVSMVERS
                     GVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSII
                     FSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREG
                     DWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDT
                     EDSSCYIPPSTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNT
                     QTASPRALADSLMQLARQVSRLESGQ"
     rep_origin      10146..10585
                     /note="Col E1"
     CDS             complement(11344..12204)
                     /codon_start=1
                     /transl_table=11
                     /product="Amp"
                     /protein_id="AAS78505.1"
                     /db_xref="GI:45861244"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     misc_feature    10013..10146
                     /note="LoxP LE"
```

```
DB:
  Query Match:      11
Best Local Similarity:      99.67%
Percent Similarity:      99.56%      Mismatches:      0      Indels:      0
Score:      1210.00      Matches:      224
Pred. No.:                  Length:      12404
Alignment Scores:
  1.24e-112
```

```
ORIGIN

Qy         1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly         20
             ::::|||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1384 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC       1325

Qy        21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly         40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1324 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC       1265

Qy        41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle         60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 21

```
LOCUS       AX370406                 678 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 3 from Patent WO0196373.
ACCESSION   AX370406
VERSION     AX370406.1  GI:18857491
KEYWORDS
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Fradkov,A.F. and Terskikh,A.
  TITLE     Fluorescent timer proteins and methods for their use
  JOURNAL   Patent: WO 0196373-A 3 20-DEC-2001;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="variant of sequence from Discosoma sp."
```

ORIGIN

```
Qy    1264 CACAACACACCGTGAAGGTGAAGGTGACCAAGGGCGGGCCCCCTGCCCTTCGCCTGGGACATC 1205
Db    1204 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1145
...
Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db     81 AspTyrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db    1144 GACTACTACAAGAAGCTGTCCTTCCCCGAGGGGTTCAAGTGGGAGAGGGTGATGAACTTCGAG 1085
...
```

Qy   221 HisHisLeuPheLeu 225
Db   724 CACCACCTGTTCCTG 710

RESULT 22

```
LOCUS       AX348043                 666 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 22 from Patent EP1172375.
ACCESSION   AX348043
VERSION     AX348043.1  GI:18614153
KEYWORDS
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Odenthal,M. and Jung,D.
  TITLE     Gene expression, genome alteration and reporter gene expression in
            myofibroblasts and myofibroblasts-like cells by using the
            regulatory regions within the alpha smooth muscle alpha-actin gene
  JOURNAL   Patent: EP 1172375-A 22 16-JAN-2002;
            Odenthal, Margarete (DE)
FEATURES             Location/Qualifiers
     source          1..666
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Red Fluorescent Protein"
```

ORIGIN

Qy   221 HisHisLeuPheLeu 225
Db   661 CACCACCTGTTCCTG 675

RESULT 23

```
Qy    US-10-006-922A-12 (1-225) x AX34043 (1-666)

      Score:                1196.00       Matches:          221
      Percent Similarity:   100.00%       Conservative:     0
      Best Local Similarity: 100.00%      Mismatches:       0
      Query Match:          98.52%        Indels:           0
DB:                                       Gaps:             0
```

```
Qy    5    LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly   24
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1    AAGAATGTTATCAAGGAGTTCATGAGGTTAAGGGTTCGCATGGAAGGAACGGTCAATGGG   60

Qy    25   HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal   44
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61   CACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCAGACCATACGAAGGCCACAATACCGTA   120

Qy    45   LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln   64
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AAGCTTAAGGTAACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCACCACAA   180

Qy    65   PheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIleProAspTyrLysLys   84
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TTTCAGTATGGAAGCAAAGTATATGTCAAGCACCCTGCCAATATACCAGACTATAAAAAG   240

Qy    85   LeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyVal   104
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAAGACGGTGGCGTC   300

Qy    105  ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe   124
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTACAAGGTCAAGTTC   360

Qy    125  IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu   144
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  ATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACAATGGGCTGGGAA   420

Qy    145  AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAla   164
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  GCCAGCACTGAGCGGTTGTATCCCCGTGATGGCGTTGTTGAAAGGAGAATTCATAAGGCT   480

Qy    165  LeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys   184
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CTGAAGCTGAAAGACGGTGGTCATTACTTGGTGGAATTCAAAAGTATTTACATGGCAAAG   540

Qy    185  LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis   204
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  AAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGATATAACAAGCCAC   600

Qy    205  AsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHisLeuPhe   224
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  AACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGCCACCATCTGTTC   660

Qy    225  Leu 225
           |||
Db    661  CTT 663
```

```
Qy    US-10-006-922A-12 (1-225) x AX353910 (1-666)
```

Alignment Scores:

```
      Pred. No.:                         1
      Score:                9.44e-113
      Percent Similarity:   100.00%       Matches:          221
      Best Local Similarity: 100.00%      Conservative:     0
      Query Match:          98.52%        Mismatches:       0
DB:                                       Indels:           0
                                          Gaps:             0
```

```
Qy    5    LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly   24
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1    AAGAATGTTATCAAGGAGTTCATGAGGTTAAGGGTTCGCATGGAAGGAACGGTCAATGGG   60

Qy    25   HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal   44
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61   CACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCAGACCATACGAAGGCCACAATACCGTA   120

Qy    45   LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln   64
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AAGCTTAAGGTAACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCACCACAA   180

Qy    65   PheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIleProAspTyrLysLys   84
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TTTCAGTATGGAAGCAAAGTATATGTCAAGCACCCTGCCAATATACCAGACTATAAAAAG   240

Qy    85   LeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyVal   104
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAAGACGGTGGCGTC   300

Qy    105  ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe   124
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTACAAGGTCAAGTTC   360

Qy    125  IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu   144
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  ATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACAATGGGCTGGGAA   420

Qy    145  AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAla   164
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  GCCAGCACTGAGCGGTTGTATCCCCGTGATGGCGTTGTTGAAAGGAGAATTCATAAGGCT   480

Qy    165  LeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys   184
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CTGAAGCTGAAAGACGGTGGTCATTACTTGGTGGAATTCAAAAGTATTTACATGGCAAAG   540

Qy    185  LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis   204
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  AAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGATATAACAAGCCAC   600

Qy    205  AsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHisLeuPhe   224
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  AACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGCCACCATCTGTTC   660

Qy    225  Leu 225
           |||
Db    661  CTT 663
```

RESULT 24

```
LOCUS       AX353910              666 bp    DNA       linear   PAT 06-FEB-2002
DEFINITION  Sequence 22 from Patent WO0204509.
ACCESSION   AX353910
VERSION     AX353910.1  GI:18618888
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Odenthal,M. and Jung,D.
  TITLE     Gene expression, genome alteration and reporter expression in
            myofibroblasts and myofibroblast-like cells
  JOURNAL   Patent: WO 0204509-A 22 17-JAN-2002;
            Odenthal, Margarete (DE)
FEATURES             Location/Qualifiers
```

RESULT 24

```
LOCUS       AY679107              711 bp    mRNA      linear   INV 25-AUG-2004
DEFINITION  Discosoma sp. RC-2004 enhanced red fluorescent protein R+ mRNA,
            complete cds.
ACCESSION   AY679107
VERSION     AY679107.1  GI:51472046
KEYWORDS    .
```

ORIGIN

```
source          1..666
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="Red Fluorescent Protein"
```

SOURCE Discosoma sp. RC-2004
ORGANISM Discosoma sp. RC-2004
  Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
  Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE 1 (bases 1 to 711)
AUTHORS Carter,R.W., Gibbs,P.D.L. and Schmale,M.C.
TITLE Cloning of Anthozoan Fluorescent Protein Genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 711)
AUTHORS Carter,R.W., Gibbs,P.D.L. and Schmale,M.C.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2004) MBF, University of Miami, 4600 Rickenbacker
  Cswy, Miami, FL 33129, USA

FEATURES Location/Qualifiers
 source 1..711
  /organism="Discosoma sp. RC-2004"
  /mol_type="mRNA"
  /db_xref="taxon:289055"
 CDS 1..711
  /note="GFP-like; DspR+"
  /codon_start=1
  /product="enhanced red fluorescent protein R+"
  /protein_id="AAU04444.1"
  /db_xref="GI:51472047"
  /translation="MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
  KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
  GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
  IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
  GRHHLPLKABLGSNVGER"

ORIGIN

Alignment Scores:
Pred. No.:   1.02e-112
Score: 1196.00
Percent Similarity: 99.11%   Length: 711
Best Local Similarity: 98.22% Matches: 221
Query Match: 98.52% Conservative: 2
DB: 2   Mismatches: 2
           Indels: 0
           Gaps: 0

US-10-006-922A-12 (1-225) x AX679107 (1-711)

```
Qy      1 MetArgSerLysAsnValIleLeuGlyGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGTTGTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCTTTGGCTTGGGATATT 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCATCCTGCCGACATACCA 240

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACTATAAAAAGCTGTCATTCCTGAAGGATTTAAATGGGAAAGGGTTCATGAACTTTGAA 300

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGTGGCGTGGTCACTGTTACTCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA 420

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||:::||||||||||||||||||||||||||||||:||||
Db    481 ATTCATAAGGCTCTGAAGTTGAAAGACGGTGGTCATTACCTTGAATTCAAAAGTATT 540

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGTTACTACTATGTTGACTCCAAACTGGAT 600

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||:::||||:||||||||||||||||||||||||||||||||||||||
Db    601 ATAACAAGGCACCACAAAAGACTATACAACATCGTTGAGCAGTATGAAAGAACCGAGGAGGAC 660

Qy    221 HisHisLeuPheLeu 225
          ||||||||||||||
Db    661 CACCACTGTTCCTT 675
```

RESULT 25

LOCUS AF545828   678 bp mRNA linear INV 04-JUN-2004
DEFINITION Discosoma sp. JW-2002 orange fluorescent protein FP586 mRNA,
  complete cds.
ACCESSION AF545828
VERSION AF545828.1 GI:33333763
KEYWORDS .
SOURCE Discosoma sp. JW-2002
ORGANISM Discosoma sp. JW-2002
  Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
  Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE 1 (bases 1 to 678)
AUTHORS Wiedenmann,J. and Girod,A.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) Abteilung Allgemeine Zoologie und
  Endokrinologie, Universitaet Ulm, Albert Einstein Allee 11, Ulm
  89069, Germany

FEATURES Location/Qualifiers
 source 1..678
  /organism="Discosoma sp. JW-2002"
  /mol_type="mRNA"
  /db_xref="taxon:208461"
  /note="isolated from specimen with smooth disc"
 CDS 1..678
  /note="GFP-like protein, orange-red; dsFP586; similar to
  fluorescent protein drFP583 (dsRed)"
  /codon_start=1
  /product="orange fluorescent protein FP586"
  /protein_id="AAQ11987.1"
  /db_xref="GI:33333764"
  /translation="MSCSKNVIKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
  KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
  GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGD
  IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
  GRHHLFL"

ORIGIN

Alignment Scores:
Pred. No.:   1.95e-112
Score: 1193.00
Percent Similarity: 98.67%   Length: 678
Best Local Similarity: 98.22% Matches: 221
Query Match: 98.27% Conservative: 1
DB: 2   Mismatches: 3
           Indels: 0
           Gaps: 0

US-10-006-922A-12 (1-225) x AF545828 (1-678)

```
Qy      1 MetArgSerLysAsnValIleLeuGlyGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGTTGTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGAGAGGGCCATACGAAGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
```

Left column (RESULT 26)

RESULT 26
AX370408

```
LOCUS       AX370408            678 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 5 from Patent WO0196373.
ACCESSION   AX370408
VERSION     AX370408.1  GI:18857492
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Fradkov,A.F. and Terskikh,A.
  TITLE     Fluorescent timer proteins and methods for their use
  JOURNAL   Patent: WO 0196373-A 5 20-DEC-2001;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="variant of sequence from Discosoma sp."
ORIGIN
```

Alignment Scores:
Pred. No.:   3.1e-112
Score: 1191.00
Percent Similarity: 98.67%
Percent Identity: 97.78%
Best Local Similarity: 98.11%
Query Match: 6

Length: 678
Matches: 220
Conservative: 2
Mismatches: 3
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AX370408 (1-678)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGluGly 20

Db  121 CACAATACCCGTAAAGGCTTAAGGTAACCAAAGGGGGGGGGACCTTGCCATTTGCTTGGGATATT 180
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCAATTTCAGTATGGAGAAGCAAGGTATATGTCAAGCACACCTGCCGACATACCA 240
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCCGTTATGCAAAAGAAGACA 420
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT 540
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTACGTGGACAGCAAACTGGAT 600
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660
Qy  221 HisHisLeuPheLeu 225
Db  661 CACCACCTGTTCCTT 675
```

Right column (RESULT 27)

RESULT 27
AX824732

```
LOCUS       AX824732            678 bp    DNA     linear   PAT 11-DEC-2003
DEFINITION  Sequence 14 from Patent WO02068459.
ACCESSION   AX824732
VERSION     AX824732.1  GI:39750595
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   .
  TITLE     Non aggregating fluorescent proteins and methods for using the same
  JOURNAL   Patent: WO 02068459-A 14 06-SEP-2002;
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="non-aggregating mutant"
ORIGIN
```

Alignment Scores:
Pred. No.:   3.1e-112
Score: 1191.00
Percent Similarity: 98.67%

Length: 678
Matches: 220
Conservative: 2

```
Db    1 ATGGGCGTTCCTCGGAGGACAGTCATCACCGGAGTTCATGCGGCTTCAAGGGTGCCATGGAGGGC 60
Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120
Qy   41 HisAsnThrValLeuValAlaThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 CTGTCCCCCCAGTTCCAGTACGGGAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTCATGAACTTCGAG 300
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660
Qy  221 HisHisLeuPheLeu 225
Db  661 CACCACCTGTTCCTG 675
```

US-10-006-922A-12 (1-225) x AX824732 (1-678)

```
Best Local Similarity:  97.78%        Mismatches:  3
Query Match:            98.11%        Indels:      0
DB:                     6             Gaps:        0
```

```
Qy    1  MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly  20
Db    1  ATGGGCTTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGTGCGCCATGGAAGGC  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db   61  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db  121  CACAACACCGTGAAGTTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db  181  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  241  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy  101  AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  301  GACGGCGGCGTGGCCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  540

Qy  181  TyrMetAlaLysArgProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  541  TACATGGCCAAGAGACCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  660

Qy  221  HisHisHisLeuPheLeu  225
Db  661  CACCACCTGTTCCTG  675
```

US-10-006-922A-12 (1-225) x AY679106 (1-921)

```
Alignment Scores:
Pred. No.:            2
Score:               4.54e-112        Length:        921
                     1191.00          Matches:       220
Percent Similarity:  98.67%           Conservative:  2
Best Local Similarity: 97.78%         Mismatches:    3
Query Match:         98.11%           Indels:        0
DB:                  2                Gaps:          0
```

```
Qy    1  MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly   20
Db   67  ATGAGTGTGTTCCAAGAATATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA  126

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db  127  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAAGGAAGGCCATACGAAGGC  186

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db  187  CACAATACCGTAAAGCTTAAGGTAACCAAAGGGGGACCTTTGCCATTTGCCTTGGGATATT  246

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db  247  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCAGCAGACATACCA  306

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  307  GACTATAAAAAGCTGTCATTTCCTGAAGGAATTAAATGGGAAAGGGTTCATGAACTTTGAA  366

Qy  101  AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  367  GACGGTGGCGTGCCTACTGTACCCCAAGATCCCAGTTTGCAAGATGGTTGTTTCATTTAC  426

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  427  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA  486

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  487  ATGGGCTGGGAAGCCAGCACAGAACGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  546

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  547  ATTCATAAGGCTCTGAAGTTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAACTATT  606

Qy  181  TyrMetAlaLysArgProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  607  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTTGAT  666
```

RESULT 29

```
LOCUS       AX824731    675 bp    DNA    linear    PAT 11-DEC-2003
DEFINITION  Sequence 13 from Patent WO02068459.
ACCESSION   AX824731
VERSION     AX824731.1  GI:39750594
KEYWORDS
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Non aggregating fluorescent proteins and methods for using the
  TITLE     same
  JOURNAL   Patent: WO 02068459-A 13 06-SEP-2002;
FEATURES             Location/Qualifiers
     source          1..675
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="non-aggregating mutant"
ORIGIN
```

US-10-006-922A-12 (1-225) x AX824731 (1-675)

Alignment Scores:
Pred. No.:

```
Score:                  1186.00          Length:       675
Percent Similarity:     98.22%           Matches:      219
Best Local Similarity:  97.33%           Conservative: 2
Query Match:            97.69%           Mismatches:   4
DB:                                      Indels:       0
                                         Gaps:         0
```

```
Qy    1 MetArgSerSerLysAsnValIleIleLysGlyGlyProPheMetArgPheLysValArgMetGluGly 20
        |||||||||||:||||||::|||||||||||||||||||||||||||||||||||||||
Db    1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATCGGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValIleLeuLysValGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAACACCGTGATCCTGAAGGTGGACGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGGCTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ...

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlnThrGluGlyArg 220
        |||||||||::|||||||||||||||||||||||||||||||||||
Db      ATAACAAGCCACAACAAAGACTATACAATCGTTGAGCAGVATAATGAAAGAACCGAGGGACGC 726

Qy  221 HisHisLeuPheLeu 225
        ||||||||||||||
Db      CACCAATCTGTTCCTT 741
```

RESULT 30

```
LOCUS       AX666133    1050 bp    DNA    linear    PAT 26-MAR-2003
DEFINITION  Sequence 7 from Patent WO02060941.
ACCESSION   AX666133
VERSION     AX666133.1  GI:29290961
KEYWORDS
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Zhao,M., Xu,M., Jiang,P. and Yang,M.
  TITLE     Fluorescent proteins
  JOURNAL   Patent: WO 02060941-A 7 08-AUG-2002;
            Anticancer, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1050
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Coral"
ORIGIN
```

US-10-006-922A-12 (1-225) x AX666133 (1-1050)

Alignment Scores:
Pred. No.:

```
Score:                  1186.00          Length:       1050
Percent Similarity:     98.22%           Matches:      219
Best Local Similarity:  97.33%           Conservative: 2
Query Match:            97.69%           Mismatches:   4
DB:                                      Indels:       0
                                         Gaps:         0
```

```
Qy    1 MetArgSerSerLysAsnValIleIleLysGlyGlyProPheMetArgPheLysValArgMetGluGly 20
        |||||||||||:||||||::||||||||||||||||||||||||||||||||||||||||
Db  289 ATGGGCTCCTCCGAGAACGTCATCACCGAGTTCATCGGCTTCAAGGTGCGCATGGAGGGC 348

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  349 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 408

Qy   41 HisAsnThrValIleLeuLysValGlyGlyProLeuProPheAlaTrpAspIle 60
        |::||||||||||||||||||||||||||||||||||||||||||||||
Db  409 CACAACACCGTGATCCTGAAGGTGGACGGCCCCCTGCCCTTCGCCTGGGACATC 468

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  469 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 528

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  529 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 588

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  589 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGGCTTCATCTAC 648

Qy  121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  649 ...
```

US-10-006-922A-12 (1-225) x CQ849509 (1-4488)

Alignment Scores:

```
Pred. No.:      1
Score:          1186.00
Percent Similarity:   98.22%
Best Local Similarity:   97.33%
Query Match:    6
DB:
```

Pred. No.:	1	
Score:	1186.00	
Percent Similarity:	98.22%	
Best Local Similarity:	97.33%	
Query Match:	6	

```
1.01e-110       Length:    4488
1186.00         Matches:   219
98.22%          Conservative:   2
97.33%          Mismatches:   4
6               Indels:    0
                Gaps:      0
```

RESULT 31

```
LOCUS        CQ849509        4488 bp    DNA        linear    PAT 23-AUG-2004
DEFINITION   Sequence 6 from Patent WO2004067751.
ACCESSION    CQ849509
VERSION      CQ849509.1  GI:51507513
KEYWORDS
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1
  AUTHORS    Couillard-Despres,S., Karl,C., Kuhn,H.G. and Aigner,L.
  TITLE      Use of regulatory sequences for specific, transient expression
             inneuronal determined cells
  JOURNAL    Patent: WO 2004067751-A 6 12-AUG-2004;
             Klinikum der Universitaet Regensburg (DE)
FEATURES             Location/Qualifiers
     source          1..4488
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="fusion construct homo sapiens and DsRed2"
ORIGIN
```

```
Qy    649 AAGGTGAAGTTCATCGGCGTGACAGTTCCCTCGGACGGCCCCGTGATGCAGAAGAAGACC 708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3812 GACTACAAGAAGCCTGTCGTCCTTCCCGGAGGGCGTTCAAGTGGGAGGAACTTCGAG 3871

Qy    709 ATGGGCTGGGAGGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGTGCTGAAGGGCGAG 768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Qy    769 ACCCACAAGGGCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGTCCATC 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Qy    829 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGCCAAGCTGGAC 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Qy    889 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCGCCGAGGGCCGC 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Qy    949 CACCACCTGTTCCTG 963
          |||||||||||||||
Db
```

RESULT 32

```
LOCUS        AY818375        4546 bp    DNA        circular SYN 12-APR-2005
DEFINITION   Cloning vector pSAT6-DsRed2-C1, complete sequence.
ACCESSION    AY818375
VERSION      AY818375.1  GI:56553574
KEYWORDS
SOURCE       Cloning vector pSAT6-DsRed2-C1
  ORGANISM   Cloning vector pSAT6-DsRed2-C1
             other sequences; artificial sequences; vectors.
REFERENCE    1  (bases 1 to 4546)
  AUTHORS    Tzfira,T., Tian,G.W., Lacroix,B., Vyas,S., Li,J., Leitner-Dagan,Y.,
             Krichevsky,A., Taylor,T., Vainstein,A. and Citovsky,V.
  TITLE      pSAT vectors: a modular series of plasmids for autofluorescent
             protein tagging and expression of multiple genes in plants
  JOURNAL    Plant Mol. Biol. 57 (4), 503-516 (2005)
  PUBMED     15821977
REFERENCE    2  (bases 1 to 4546)
  AUTHORS    Tzfira,T., Lacroix,B. and Citovsky,V.
  TITLE      Direct Submission
  JOURNAL    Submitted (05-NOV-2004) Department of Biochemistry and Cell
             Biology, State University of New York Stony Brook, 100 Nicolls Rd,
             Stony Brook, NY 11794, USA
FEATURES             Location/Qualifiers
     source          1..4546
                     /organism="Cloning vector pSAT6-DsRed2-C1"
                     /mol_type="other DNA"
                     /db_xref="taxon:301574"
     promoter        441..762
                     /note="CaMV 35S promoter"
     promoter        768..1089
                     /note="CaMV 35S promoter"
     misc_feature    1190..1320
                     /note="5' UTR from tobacco etch virus; translational
                     enhancer"
     misc_feature    1323..1997
                     /note="N-terminal end of DsRed2 protein; part of fusion
                     protein"
     misc_feature    1998..2066
                     /note="multiple cloning site; MCS"
     terminator       2067..2277
```

```
Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   3812

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   3872 GACGGCGGCGTGGTGACCGTGACCCAGGACTCTCCCTGCAGGACGGCGTGCTTCATCTAC 3931

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   3932 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 3991

Qy   141 MetGlyTrpGluAlaSerThrGluArgMetTyrProGluAspGlyValLeuLysGlyGlu 160
Db   3992 ATGGGCTGGGAGGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGTGCTGAAGGGCGAG 4051

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   4052 ATCCACAAGGGCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGTCCATC 4111

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   4112 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 4171

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   4172 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 4231

Qy   221 HisHisLeuPheLeu 225
Db   4232 CACCACCTGTTCCTG 4246
```

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3572 ATGGGCTCCTCCAGAAGGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3631

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3632 ACCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3691

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3692 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3751

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3752 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3811
```

Left column

```
CDS           /note="CaMV 35S terminator"
              complement(3486..4346)
              /note="amp resistance"
              /codon_start=1
              /product="bla"
              /protein_id="AAV97912.1"
              /db_xref="GI:56553575"
              /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
              IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
              YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
              DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
              LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVIYTTGSQATMDERNRQIA
              EIGASLIKHW"

ORIGIN
Alignment Scores:
Pred. No.:        1.03e-110
Score:            1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match:        97.69%
DB:                 11

                  Length:      4546
                  Matches:     219
                  Conservative: 2
                  Mismatches:  4
                  Indels:      0
                  Gaps:        0

US-10-006-922A-12 (1-225) x AY818375 (1-4546)

Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1323 ATGGCCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1382

Qy    21 ThrValAsnGlyHisGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1383 ACCGTGAACGGCCACGGCTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1442

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1443 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1502

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1503 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1562

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1563 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1622

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1623 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1682

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1683 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1742

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1743 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1802

Qy   161 IleLysHisLysLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1803 ACCAAGCACAAGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1862

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db  1863 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 1922

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1923 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 1982

Qy   221 HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db  1983 CACCACCTGTTCCTG 1997
```

Right column

```
RESULT 33

LOCUS       DQ005468               4555 bp   DNA     circular SYN 07-MAY-2005
DEFINITION  Cloning vector pSAT6A-DsRed2-N1, complete sequence.
ACCESSION   DQ005468
VERSION     DQ005468.1  GI:63002530
KEYWORDS    .
SOURCE      Cloning vector pSAT6A-DsRed2-N1
  ORGANISM  Cloning vector pSAT6A-DsRed2-N1
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4555)
  AUTHORS   Chung,S.-M. and Tzfira,T.
  TITLE     A versatile vector system for multiple gene expression in plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4555)
  AUTHORS   Chung,S.-M., Vyas,S. and Tzfira,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2005) Department of Biochemistry and Cell
            Biology, State University of New York at Stony Brook, Stony Brook,
            NY 11794, USA
FEATURES             Location/Qualifiers
     source          1..4555
                     /organism="Cloning vector pSAT6A-DsRed2-N1"
                     /mol_type="other DNA"
                     /db_xref="taxon:326511"
     promoter        441..762
                     /note="CaMV 35S promoter"
     promoter        768..1089
                     /note="CaMV 35S promoter"
     enhancer        1190..1320
                     /note="translational enhancer 5'-UTR from tobacco etch
                     virus"
     misc_feature    1321..1379
                     /note="multiple cloning site; MCS"
     CDS             1380..2057
                     /codon_start=1
                     /product="DsRed2"
                     /protein_id="AAY25372.1"
                     /db_xref="GI:63002531"
                     /translation="MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVATVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGB
                     THRALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYIVEQYERTB
                     GRHLFLL"
     terminator      2076..2286
                     /note="CaMV 35S terminator"
     CDS             complement(3495..4355)
                     /note="amp resistance"
                     /codon_start=1
                     /product="Bla"
                     /protein_id="AAY25373.1"
                     /db_xref="GI:63002532"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"

ORIGIN

Alignment Scores:
Pred. No.:        1.03e-110
Score:            1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match:        97.69%
DB:                 11

                  Length:      4555
                  Matches:     219
                  Conservative: 2
                  Mismatches:  4
                  Indels:      0
                  Gaps:        0

US-10-006-922A-12 (1-225) x DQ005468 (1-4555)

Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1380 ATGGCCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1439
```

```
Qy    21  ThrValAsnGlyHisGluPheGlnIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  1440  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1499

Qy        HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle
Db  1500  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1559

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  1560  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1619

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  1620  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1679

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  1560  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1679

Qy   141  MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  1800  ATGGGCTACGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1859

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  1860  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1919

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  1920  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 1979

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  1980  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2039

Qy   221  HisHisHisLeuPheLeu 225
Db  2040  CACCACCACCTGTTCCTG 2054
```

```
RESULT 34
AY818373
LOCUS        AY818373                4570 bp    DNA     circular SYN 12-APR-2005
DEFINITION   Cloning vector pSAT6-DsRed2-N1, complete sequence.
ACCESSION    AY818373
VERSION      AY818373.1  GI:56553569
KEYWORDS     .
SOURCE       Cloning vector pSAT6-DsRed2-N1
  ORGANISM   Cloning vector pSAT6-DsRed2-N1
             other sequences; artificial sequences; vectors.
REFERENCE    1  (bases 1 to 4570)
  AUTHORS    Tzfira,T., Tian,G.W., Lacroix,B., Vyas,S., Li,J., Leitner-Dagan,Y.,
             Krichevsky,A., Taylor,T., Vainstein,A. and Citovsky,V.
  TITLE      pSAT vectors: a modular series of plasmids for autofluorescent
             protein tagging and expression of multiple genes in plants
  JOURNAL    Plant Mol. Biol. 57 (4), 503-516 (2005)
  PUBMED     15821977
REFERENCE    2  (bases 1 to 4570)
  AUTHORS    Tzfira,T., Lacroix,B. and Citovsky,V.
  TITLE      Direct Submission
  JOURNAL    Submitted (05-NOV-2004) Department of Biochemistry and Cell
             Biology, State University of New York Stony Brook, 100 Nicolls Rd,
             Stony Brook, NY 11794, USA
FEATURES             Location/Qualifiers
     source          1..4570
                     /organism="Cloning vector pSAT6-DsRed2-N1"
                     /mol_type="other DNA"
                     /db_xref="taxon:301572"
```

```
     promoter        441..762
                     /note="CaMV 35S promoter"
     promoter        768..1090
                     /note="CaMV 35S promoter"
     misc_feature    1190..1320
                     /note="CaMV 35S promoter"
     misc_feature    1323..1394
                     /note="5' UTR from tobacco etch virus; translational
                     enhancer"
     CDS             1395..2072
                     /note="multiple cloning site; MCS"
                     /codon_start=1
                     /product="DsRed2"
                     /protein_id="AAV97910.1"
                     /db_xref="GI:56553571"
                     /translation="MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVATVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     THKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
     terminator      2091..2301
                     /note="CaMV 35S terminator"
     CDS             complement(3510..4370)
                     /note="amp resistance"
                     /codon_start=1
                     /product="bla"
                     /protein_id="AAV97909.1"
                     /db_xref="GI:56553570"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
```

```
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                    1.04e-110
Percent Similarity:       1186.00
Best Local Similarity:    98.22%
Query Match:              97.33%
DB:                       97.69%
                          11
```

```
US-10-006-922A-12 (1-225) x AY818373 (1-4570)
```

```
Length:         4570
Matches:        219
Conservative:   2
Mismatches:     4
Indels:         0
Gaps:           0
```

```
Qy     1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db  1395  ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1454

Qy    21  ThrValAsnGlyHisGluPheGlnIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  1455  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1514

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  1515  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1574

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  1575  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1634

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  1635  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1694

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  1695  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1754

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  1755  AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 1814
```

```
Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuGlyGlyGlu 160
     141 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTGCTGCTGGGAGGGGAG 1874
Db  1815 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTGCTGCTGGGAGGGGAG 1874

Qy   161 IleHisLysAlaLeuLeuLysGlyHisSerTyrLeuValGluPheLysSerIle 180
     161 ACCCACAAGGCCCTGAAGCTTGAAGGACACGGGGGGGGGCGGGGGGGGCGG 180
Db  1875 ACCCACAAGGCCCTGAAGCTTGAAGGACACGGGGGGGGGCGGGGGGGGCGG 1934

Qy   181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuGlyAsp 200
     181 TACATGGCCAAGGAAGGCCCGTGCAGTTGCCCGGCTACTACTACGTGGACAGCAAGCTGGGAC 1994
Db  1935 TACATGGCCAAGGAAGGCCCGTGCAGTTGCCCGGCTACTACTACGTGGACAGCAAGCTGGGAC 1994

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
     201 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2054
Db  1995 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2054

Qy   221 HisHisSerLeuPheLeu 225
     221 CACCACCTGTTCCTG 2069
Db  2055 CACCACCTGTTCCTG 2069
```

```
RESULT 35
AY640628/c

LOCUS       AY640628                4829 bp    DNA     circular SYN 01-JUL-2005
DEFINITION  siRNA vector pSUPER-hSyn-DsRed2N1-CytB-AS, complete sequence.
ACCESSION   AY640628
VERSION     AY640628.1  GI:56119185
KEYWORDS    .
SOURCE      siRNA vector pSUPER-hSyn-DsRed2N1-CytB-AS
  ORGANISM  siRNA vector pSUPER-hSyn-DsRed2N1-CytB-AS
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4829)
  AUTHORS   Michel,U., Malik,I., Ebert,S., Bahr,M. and Kugler,S.
  TITLE     Long-term in vivo and in vitro AAV-2-mediated RNA interference in
            rat retinal ganglion cells and cultured primary neurons
  JOURNAL   Biochem. Biophys. Res. Commun. 326 (2), 307-312 (2005)
  PUBMED    15582578
REFERENCE   2  (bases 1 to 4829)
  AUTHORS   Michel,U. and Kugler,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2004) Neurology, University of Goettingen,
            Waldweg 33, Goettingen 37073, Germany
FEATURES             Location/Qualifiers
     source          1..4829
                     /organism="siRNA vector pSUPER-hSyn-DsRed2N1-CytB-AS"
                     /mol_type="other DNA"
                     /db_xref="taxon:297630"
     promoter        135..645
                     /note="R1-Ori"
     rep_origin      625..645
                     /note="T7 promoter"
     promoter        739..899
                     /note="SV40-pA; SV40 polyadenylation sequence"
     polyA_signal    899..1039
                     /note="SV40 chimeric intron derived from pCI-NEO"
     intron          1045..1722
                     /note="SV40 chimeric intron derived from pCI-NEO"
     complement(1045..1722)
                     /gene="DsRed2"
     CDS             complement(1045..1722)
                     /gene="DsRed2"
                     /note="derived from Clontech's pDsRed2-N1"
                     /codon_start=1
                     /transl_table=1
                     /product="DsRed2"
                     /protein_id="AAV73958.1"
                     /db_xref="GI:56119186"
                     /translation="MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVARVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     THKALKLKDGGHYIVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
     misc_feature    1741..1759
                     /note="MCS; multiple cloning site"
     promoter        complement(1759..2228)
                     /note="HSYN-promoter; human synapsin 1 gene promoter (-422
```

```
                     to +53)"
     promoter        2240..2455
                     /note="disabled human H1 RNA promoter"
     misc_feature    2456..2587
                     /note="CytB-AS; partial human cytochrome B antisense
                     sequence"
     promoter        complement(2640..2658)
                     /note="T3 promoter"
     rep_origin      3026..3693
                     /note="pUC Ori"
     gene            complement(3841..4701)
                     /gene="Amp"
     CDS             complement(3841..4701)
                     /gene="Amp"
                     /note="ampicillin resistance"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAV73959.1"
                     /db_xref="GI:56119187"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"

ORIGIN
```

```
US-10-006-922A-12 (1-225) x AY640628 (1-4829)

Alignment Scores:
Pred. No.:            1.11e-110
Score:               1186.00
Percent Similarity:  98.22%
Best Local Similarity: 97.33%
Query Match:         97.69%
DB:                  11
```

	Length:	4829
	Matches:	219
	Conservative:	2
	Mismatches:	4
	Indels:	0
	Gaps:	0

```
Qy     1 MetArgSerSerLysAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
       1 ATGGCCTCCACCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1663
Db  1722 ATGGCCTCCACCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1663

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
      21 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1603
Db  1662 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1603

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
      41 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1543
Db  1602 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1543

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
      61 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1483
Db  1542 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1483

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
      81 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1423
Db  1482 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1423

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
     101 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1363
Db  1422 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
     121 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1303
Db  1362 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1303

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuGlyGlu 160
     141 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTGCTGCTGGGAGGGGAG 160
Db  1302 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTGCTGCTGGGAGGGGAG 1243

Qy   161 IleHisLysAlaLeuLeuLysGlyHisSerTyrLeuValGluPheLysSerIle 180
     161 ACCCACAAGGCCCTGAAGCTTGAAGGACACGGGGGGGGGCGGGGGGGGCGG 180
Db  1242 ACCCACAAGGCCCTGAAGCTTGAAGGACACGGGGGGGGGCGGGGGGGGCGG 1183
```

RESULT 36

```
LOCUS       AY640625/c               4906 bp    DNA     circular SYN 01-JUL-2005
DEFINITION  siRNA vector pSUPER-CMV-DsRed2N1-CytB-AS, complete sequence.
ACCESSION   AY640625
VERSION     AY640625.1  GI:56119176
KEYWORDS    .
SOURCE      siRNA vector pSUPER-CMV-DsRed2N1-CytB-AS
  ORGANISM  siRNA vector pSUPER-CMV-DsRed2N1-CytB-AS
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4906)
  AUTHORS   Michel,U., Malik,I., Ebert,S., Bahr,M. and Kugler,S.
  TITLE     Long-term in vivo and in vitro AAV-2-mediated RNA interference in
            rat retinal ganglion cells and cultured primary neurons
  JOURNAL   Biochem. Biophys. Res. Commun. 326 (2), 307-312 (2005)
  PUBMED    15582578
REFERENCE   2  (bases 1 to 4906)
  AUTHORS   Michel,U. and Kugler,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2004) Neurology, University of Goettingen,
            Waldweg 33, Goettingen 37073, Germany
FEATURES             Location/Qualifiers
     source          1..4906
                     /organism="siRNA vector pSUPER-CMV-DsRed2N1-CytB-AS"
                     /mol_type="other DNA"
                     /db_xref="taxon:297627"
     rep_origin      135..441
                     /note="F1-Ori"
     promoter        625..645
                     /note="T7 promoter"
     polyA_signal    complement(739..899)
                     /note="SV40-pA; SV40 polyadenylation sequence"
     intron          complement(899..1039)
                     /note="SV40 chimeric intron derived from pCI-NEO"
     gene            complement(1045..1722)
                     /gene="DsRed2"
                     /note="derived from Clontech's pDsRed2-N1"
     CDS             complement(1045..1722)
                     /gene="DsRed2"
                     /codon_start=1
                     /transl_table=11
                     /product="DsRed2"
                     /protein_id="AAV73952.1"
                     /db_xref="GI:56119177"
                     /translation="MASSENVITEFMRFKVRMEGTVNGHPEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVATVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     THKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
     promoter        complement(1741..1760)
                     /note="MCS'; multiple cloning site"
     promoter        complement(1771..2293)
                     /note="MCMV promoter; mcmv promoter (-491 to +36)"
     misc_feature    2294..2309
                     /note="MCS'; multiple cloning site"
     promoter        2317..2532
                     /note="disabled human H1 RNA promoter"
     misc_feature    2533..2664
                     /note="Cytb-AS; partial human cytochrome B antisense
                     sequence"
     promoter        complement(2717..2735)
                     /note="T3 promoter"
     rep_origin      3103..3770
                     /note="pUC Ori"
     gene            complement(3918..4778)
                     /gene="Amp"
                     /note="ampicillin resistance"
     CDS             complement(3918..4778)
                     /gene="Amp"
                     /note="ampicillin resistance"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAV73953.1"
                     /db_xref="GI:56119178"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLITDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVIYITGSQATMDERNRQIA
                     EIGASLIKHW"
ORIGIN
```

US-10-006-922A-12 (1-225) x AY640625 (1-4906)

Alignment Scores:

Pred. No.:	1	Length:	4906
Score:	1186.00	Matches:	219.
Percent Similarity:	98.22%	Conservative:	2
Best Local Similarity:	97.33%	Mismatches:	4
Query Match:	11	Indels:	0
		Gaps:	0

```
Qy     1 MetArgSerLysSerAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1722 ATGCGCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1663

Qy    21 ThrValAsnGlyHisProGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly  40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1662 ACCGTGAACGGCCACCCAGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCGGC 1603

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1602 CACAACACCGTGAAGCTGAAGGTCACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1543

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1542 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1483

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1482 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1423

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1422 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1362 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1303

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1302 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1243

Qy   161 ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1242 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1183

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1182 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGCCAAGCTGGAC 1123

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||::::|||||||||||||||||||||||||||||||||||
Db  1122 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 1063

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db  1062 CACCACCTGTTCCTG 1048
```

RESULT 37

```
Db 1122 ATCACCTCCCACAACGGAGGACTACACCATCGTGGAGCAGTACGAGCGGCCACCGAGGGGCGC 1063
        ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Qy  221 HisHisLeuPheLeu 225
Db 1062 CACCACCTGTTCCTG 1048
```

```
AJ851284
LOCUS       AJ851284                5311 bp    DNA     circular SYN 12-NOV-2004
DEFINITION  Cloning vector pRU1104 dsRed2 gene for Ds Red fluorescent protein
            2.
ACCESSION   AJ851284
VERSION     AJ851284.1  GI:55724874
KEYWORDS    Cloning vector pRU1104
SOURCE      Cloning vector pRU1104
  ORGANISM  Cloning vector pRU1104
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE     High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5311)
  AUTHORS   Poole,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..5311
                     /organism="Cloning vector pRU1104"
                     /mol_type="other DNA"
                     /lab_host="Escherichia coli"
                     /db_xref="taxon:299179"
     gene            326..1003
                     /gene="dsRed2"
     CDS             326..1003
                     /gene="dsRed2"
                     /codon_start=1
                     /transl_table=11
                     /product="Ds Red fluorescent protein 2"
                     /protein_id="CAH64889.1"
                     /db_xref="GI:55724875"
                     /translation="MASSENVITEFMRFKVMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVARVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     THKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVSQYERTE
                     GRHHLFL"
ORIGIN
```

Alignment Scores:

Pred. No.:	1
Score:	1.25e-110
Percent Similarity:	1186.00
Best Local Similarity:	98.22%
Query Match:	97.33%
DB:	97.69%
	11

Length:	5311
Matches:	219
Conservative:	2
Mismatches:	4
Indels:	0
Gaps:	0

US-10-006-922a-12 (1-225) x AJ851284 (1-5311)

```
Qy    1 MetArgSerLysAsnValIleLeuLysGlyProPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  326 ATGGGCCTCCTGCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCCATGGAGGGC 385
Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  386 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 445
Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  446 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 505
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 38

```
AY613997
Db  986 CACCACCTGTTCCTG 1000
        ||||||||||||||
Qy  221 HisHisLeuPheLeu 225
```

```
LOCUS       AY613997                6423 bp    DNA     circular SYN 26-JUN-2004
DEFINITION  Cloning vector pSRalphaneoR, complete sequence.
ACCESSION   AY613997
VERSION     AY613997.1  GI:48995616
KEYWORDS    .
SOURCE      Cloning vector pSRalphaneoR
  ORGANISM  Cloning vector pSRalphaneoR
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 6423)
  AUTHORS   Allard,V., de Leseleuc,L. and Denis,F.
  TITLE     A family of mammalian expression vectors with different selection
            markers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6423)
  AUTHORS   Allard,V., de Leseleuc,L. and Denis,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2004) INRS-Institut Armand-Frappier, 531
            Boulevard des Prairies, Laval, Quebec H7V 1B7, Canada
FEATURES             Location/Qualifiers
     source          1..6423
                     /organism="Cloning vector pSRalphaneoR"
                     /mol_type="other DNA"
                     /db_xref="taxon:279783"
     misc_feature    100..686
                     /note="ECMV internal ribosomal entry site"
     misc_feature    786..1360
                     /note="ECMV internal ribosomal entry site"
     gene            1431..2108
                     /gene="dsRED2"
     CDS             1431..2108
                     /gene="dsRED2"
                     /codon_start=1
                     /transl_table=11
                     /product="red fluorescent protein"
                     /protein_id="AAT48428.1"
                     /db_xref="GI:48995617"
                     /translation="MASSENVITEFMRFKVMEGTVNGHEFEIEGEGEGRPYEGHNTV
```

```
Db  506 CTGTCCCCCCAGTTCCAGTACGGGCTCCAAGGTGTACGTGAAGCACCCCGCGGACATCCCC 565
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  566 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 625
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  626 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 685
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  686 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 745
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  746 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 805
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  806 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 865
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaLysLeuAsp 200
Db  866 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGCCAAGCTGGAC 925
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValSerGlnTyrGluArgThrGluGlyArg 220
Db  926 ATCACCTCCCACAACGAGGACTACACCATCGTGAGCCAGTACGAGCGGACCGAGGGCCGC 985
```

```
misc_feature    2134..2266
                /note="contains SV40 polyA signal"
gene            complement(2294..3088)
                /gene="aphII"
CDS             complement(2294..3088)
                /gene="aphII"
                /codon_start=1
                /transl_table=11
                /product="neomycin phosphotransferase"
                /protein_id="AAT48429.1"
                /db_xref="GI:4899561#"
                /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                PVLFVKTDLSGALNELQDEAARLSWLATGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
promoter        complement(3456..3783)
                /note="SV40 early"
gene            3914..4774
                /gene="bla"
CDS             3914..4774
                /gene="bla"
                /codon_start=1
                /transl_table=1
                /product="ampicillin resistance protein"
                /protein_id="AAT48430.1"
                /db_xref="GI:4899561#"
                /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVE
                YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                DRWEPELNEAIPNDERDTTMPAAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                EIGASLIKHW"
rep_origin      4961..5535
                /note="ColE1"
promoter        5786..6408
                /note="SRalpha"
```

```
ORIGIN
US-10-006-922A-12 (1-225) x AY613997 (1-6423)

Alignment Scores:
Pred. No.:
Score:              1.57e-110
Score:              1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match:        97.69%
DB:                 11

1   MetArgSerLysAsnValIleLysGlyValMetGluGly    20
Qy  ||| ||||||||||||||||||||||||||||||||||
Db  ATGGCCTCTCCGAGAAGGTCATCACCGAGTTCATGGCCGGGGGC 1490

21  ThrValAsnGlyHisGlyPheGluIleGluGlyGluGlyArgProTyrGluGly  40
Qy  ||||||||||||||||||||||||||||||||||||||||||||||
Db  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCCGCCCCTACGAGGGC 1550

41  HisAsnThrValIleLysGlyValThrIleLysGlyValTyrGlyProLeuProPheAlaTrpAspIle  60
Qy  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  CACAACACCGTGAAGGCTGAAGGTGAAGGTCTACGTGGGCCCCCTGCCCTTCGCCTGGGACATC 1610

61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Qy  ||||||||||||||||||||||||||||||||||||||||||||
Db  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCAGACATCCC 1670

81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Qy  ||||| |||||||||||||||||||||||||||||||||||||
Db  GACTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1730
```

```
Qy  101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  GACGGCGGCGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1790

Qy  121 LysValLysPheIleGlyValAsnProSerAspGlyProValMetGlnLysLysThr  140
Db  AAGGTGAAGTTCATCGGCGTGAACCTTCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1850

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1910

Qy  161 IleHisHisLysPheAlaIleLeuLysGluAspGlyGlyValHisTyrLeuValGlnTyrTyrGlyValSerIle  180
Db  ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCGTCCACTACCTGGTGGAGTTCAAGTCCATC 1970

Qy  181 TyrMetAlaLysGlyLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  TACATGGCCAAGGGCCCTGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 2030

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlnThrGlyGlyArg  220
Db  ATCACCTCCCACAACGAGGACTACACCATCGTGCAGGAGACGGGCGGCAGG 2090

Qy  221 HisHisLeuPheLeu  225
Db  CACCACCTGTTCCTG 2105
```

```
RESULT 39
AY640634/c
LOCUS       AY640634                6990 bp  DNA  circular SYN 01-JUL-2005
DEFINITION  siRNA vector pAAV9(5)-hSyn-DsRed2N1-EGFP-siRNA, complete sequence.
ACCESSION   AY640634
VERSION     AY640634.1  GI:56119203
KEYWORDS    .
SOURCE      siRNA vector pAAV9(5)-hSyn-DsRed2N1-EGFP-siRNA
  ORGANISM  siRNA vector pAAV9(5)-hSyn-DsRed2N1-EGFP-siRNA
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 6990)
  AUTHORS   Michel,U., Malik,I., Ebert,S., Bahr,M. and Kugler,S.
  TITLE     Long-term in vivo and in vitro AAV-2-mediated RNA interference in
            rat retinal ganglion cells and cultured primary neurons
  JOURNAL   Biochem. Biophys. Res. Commun. 326 (2), 307-312 (2005)
  PUBMED    15582578
REFERENCE   2  (bases 1 to 6990)
  AUTHORS   Michel,U., Malik,I. and Kugler,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2004) Neurology, University of Goettingen,
            Waldweg 33, Goettingen 37073, Germany
FEATURES             Location/Qualifiers
     source          1..6990
                     /organism="siRNA vector pAAV9(5)-hSyn-DsRed2N1-EGFP-siRNA"
                     /mol_type="other DNA"
                     /db_xref="taxon:297636"
     polyA_signal    complement(38..193)
                     /note="SV40-pA; SV40 polyadenylation sequence"
     intron          complement(198..341)
                     /note="SV40 chimeric intron derived from pCI-NEO"
     gene            complement(348..1025)
                     /gene="DsRed2"
                     /note="derived from Clontech's pDsRed2-N1"
     CDS             complement(348..1025)
                     /gene="DsRed2"
                     /codon_start=1
                     /transl_table=11
                     /product="DsRed2"
                     /protein_id="AAV73970.1"
                     /db_xref="GI:56119204"
                     /translation="MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVATVTQDSSLQDGCFIYKVKFIGVNPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     THRALKLKDGGHYLVEFKSIYMAKKVQLPGYYYVDAKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
```

```
misc_feature    complement(1044..1062)
                /note="MCS; multiple cloning site"
promoter        complement(1062..1531)
                /note="H1SYN-promoter; human synapsin 1 gene promoter (-422
                to +53)"
promoter        1543..1758
                /note="human H1 RNA promoter"
misc_feature    1759..1822
                /note="siEGFP; 64 bp siRNA sequence"
gene            complement(1827..4090)
                /gene="9(5)"
                /note="partial sequence from the non-coding porcine RNA UM
                9(5)"
repeat_region   4110..4250
                /note="right ITR"
                /rpt_family="inverted"
gene            5167..6027
                /gene="Amp"
CDS             5167..6027
                /gene="Amp"
                /note="ampicillin resistance"
                /codon_start=1
                /transl_table=11
                /product="beta-lactamase"
                /protein_id="AAV73971.1"
                /db_xref="GI:56119205"
                /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRIHYSQNDLVE
                YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                EIGASLIKHW"
repeat_region   6848..6988
                /note="left ITR"
                /rpt_family="inverted"
ORIGIN
```

Alignment Scores:
```
Pred. No.:              1.74e-110      Length:          6990
Score:                  1186.00        Matches:         219
Percent Similarity:     98.22%         Conservative:    2
Best Local Similarity:  97.33%         Mismatches:      4
Query Match:            97.69%         Indels:          0
DB:                     11             Gaps:            0
```

US-10-006-922A-12 (1-225) x AY640634 (1-6990)

```
Qy    1 MetArgSerLysAsnValIleLeuGlyMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db 1025 ATGCGCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGGAGGC 966

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  965 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 906

Qy   41 HisAsnThrValLeuLeuLysGlyGlyValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  905 CACAACACCGTGAACCTGAAGGGCGGCGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 846

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  845 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 786

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  785 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 726

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  725 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 666

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  665 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 606

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  605 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 546

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  545 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 486

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||:||||||||||||:|||||||||||||||
Db  485 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 426

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  425 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 366

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  365 CACCACCTGTTCCTG 351
```

RESULT 40
AY640633/c
```
LOCUS       AY640633                7058 bp    DNA     circular SYN 01-JUL-2005
DEFINITION  siRNA vector pAAV9(5)-hSyn-DsRed2N1-CytB-AS-ohneNot, complete
            sequence.
ACCESSION   AY640633
VERSION     AY640633.1  GI:56119200
KEYWORDS    .
SOURCE      siRNA vector pAAV9(5)-hSyn-DsRed2N1-CytB-AS-ohneNot
  ORGANISM  siRNA vector pAAV9(5)-hSyn-DsRed2N1-CytB-AS-ohneNot
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 7058)
  AUTHORS   Michel,U., Malik,I., Ebert,S., Bahr,M. and Kugler,S.
  TITLE     Long-term in vivo and in vitro AAV-2-mediated RNA interference in
            rat retinal ganglion cells and cultured primary neurons
  JOURNAL   Biochem. Biophys. Res. Commun. 326 (2), 307-312 (2005)
   PUBMED   15582578
REFERENCE   2  (bases 1 to 7058)
  AUTHORS   Michel,U., Malik,I. and Kugler,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2004) Neurology, University of Goettingen,
            Waldweg 33, Goettingen 37073, Germany
FEATURES             Location/Qualifiers
     source          1..7058
                     /organism="siRNA vector
                     pAAV9(5)-hSyn-DsRed2N1-CytB-AS-ohneNot"
                     /mol_type="other DNA"
                     /db_xref="taxon:297635"
     repeat_region   1..141
                     /note="left ITR"
                     /rpt_family="inverted"
     polyA_signal    complement(38..193)
                     /note="SV40-pA"; SV40 polyadenylation sequence"
     intron          complement(198..341)
                     /note="SV40 chimeric intron derived from pCI-NEO"
     gene            complement(491..1168)
                     /gene="DsRed2"
     CDS             complement(491..1168)
                     /gene="DsRed2"
                     /note="derived from Clontech's pDsRed2-N1"
                     /codon_start=1
                     /transl_table=11
                     /product="DsRed2"
                     /protein_id="AAV73969.1"
                     /db_xref="GI:56119202"
                     /translation="MASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVATVTQDSSLQDGCFIYKVKFIGVNFPSDGYVWQKKTMGWEASTERLYPRDGVLKGE
                     THKALKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
     misc_feature    complement(1187..1205)
```

promoter   /note="MCS; multiple cloning site"
                  complement(1205..1674)
                  /note="HSYN-promoter; human synapsin 1 gene promoter (-422
                  to +53)"

promoter   1686..1901
                  /note="disabled human H1 RNA promoter"

misc_feature   1902..2033
                  /note="Cytb-AS; partial human cytochrome B antisense
                  sequence"

promoter   complement(2038..4301)
                  /note="9(5)"
                  /gene="9(5)"
                  /note="partial sequence from the non-coding porcine RNA UM
                  9(5)"

repeat_region   4321..4461
                  /note="right ITR"
                  /rpt_family="inverted"

gene   5378..6238
                  /gene="Amp"

CDS   5378..6238
                  /gene="Amp"
                  /note="ampicillin resistance"
                  /codon_start=1
                  /transl_table=11
                  /product="beta-lactamase"
                  /protein_id="AAV73968.1"
                  /db_xref="GI:56119201"
                  /translation="MSIQHFRVALIPFFAAFCLPVPAHPETLVKVKDAEDQLGARVGY
                  IELDLNSGKILESFRPEERFPMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                  YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                  DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                  LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVIYTTGSQATMDERNRQIA
                  EIGASLIKHW"

ORIGIN

Alignment Scores:
Pred. No.:   1.76e-110   Length: 7058
Score: 1186.00   Matches: 219
Percent Similarity: 98.22%   Conservative: 2
Best Local Similarity: 97.33%   Mismatches: 4
Query Match: 97.69%   Indels: 0
DB: 11   Gaps: 0

US-10-006-922A-12 (1-225) x AY640633 (1-7058)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1168 ATGGGCCTCCTCCGAGAACGTCATCACCGAGTTCATCGCGCTTCAAGGTGCGATGGAGGGC 1109

Qy   21 ThrValAsnGlyHisGlyLeuProGlyGlyGlyGlyGlyGlyGlyArgProTyrGluGlyGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1108 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCCGAGGGCCGAGGGCCCCCTACGAGGGC 1049

Qy   41 HisAsnThrValIleLysThrLysLeuLysLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1048 CACAACACCGTGAAGCTGAAGAAAGGTGACCCAAGGCGGGGCCCCCTGCCCTTCGCCTGGGACATC 989

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 988 CTGTCCCCGCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCAACATCCCC 929

Qy   81 AspThrLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 928 GACACAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 869

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 868 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 809

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 808 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 749

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 748 ATGGGCTGGGAGGCTTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 689

Qy 161 IleHisLysLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 688 ACCCACAAGCCGCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 629

Qy 181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db 628 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 569

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 568 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 509

Qy 221 HisHisLeuPheLeu 225
          |||||||||||||||
Db 508 CACCACCTGTTCCTG 494

RESULT 41

LOCUS   AY640630   7147 bp   DNA   circular SYN 01-JUL-2005
DEFINITION SiRNA vector pAAV9(5)-CMV-DsRed2N1-CytB-AS-ohneNot, complete
            sequence.
ACCESSION   AY640630
VERSION   AY640630.1 GI:56119191
KEYWORDS   .
SOURCE   siRNA vector pAAV9(5)-CMV-DsRed2N1-CytB-AS-ohneNot
  ORGANISM siRNA vector pAAV9(5)-CMV-DsRed2N1-CytB-AS-ohneNot
            other sequences; artificial sequences; vectors.
REFERENCE   1 (bases 1 to 7147)
  AUTHORS   Michel,U., Malik,I., Ebert,S., Bahr,M. and Kugler,S.
  TITLE   Long-term in vivo and in vitro AAV-2-mediated RNA interference in
            rat retinal ganglion cells and cultured primary neurons
  JOURNAL   Biochem. Biophys. Res. Commun. 326 (2), 307-312 (2005)
  PUBMED   15582578
REFERENCE   2 (bases 1 to 7147)
  AUTHORS   Michel,U., Malik,I. and Kuegler,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-JUN-2004) Neurology, University of Goettingen,
            Waldweg 33, Goettingen 37073, Germany
FEATURES   Location/Qualifiers
     source   1..7147
                     /organism="siRNA vector
                     pAAV9(5)-CMV-DsRed2N1-CytB-AS-ohneNot"
                     /mol_type="other DNA"
                     /db_xref="taxon:297632"

     repeat_region   1..141
                     /note="left ITR"
                     /rpt_family="inverted"

     polyA_signal   complement(181..338)
                     /note="SV40-pA; SV40 polyadenylation sequence"

     intron   complement(341..481)
                     /note="SV40 chimeric intron derived from pCI-NEO"

     gene   complement(491..1168)
                     /gene="DsRed2"

     CDS   complement(491..1168)
                     /gene="DsRed2"
                     /note="derived from Clontech's pDsRed2-N1"
                     /codon_start=1
                     /transl_table=11
                     /product="DsRed2"
                     /db_xref="GI:56119193"
                     /protein_id="AAV73963.1"
                     /translation="MASSENVITEPMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTV
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVATVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
                     THKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"

     misc_feature   complement(1187..1206)
                     /note="MCS; multiple cloning site"

     promoter   complement(1217..1739)
                     /note="MCMV promoter; mcmv promoter (-491 to +36)"

```
promoter        1775..1990
                /note="disabled human H1 RNA promoter"
misc_feature    1991..2122
                /note="Cytb-AS; partial human cytochrome B antisense
                sequence"
                complement(2127..4390)
gene            /gene="9(5)"
                complement(2127..4390)
                /gene="9(5)"
                /note="partial sequence from the non-coding porcine RNA UM
                9(5)"
                4410..4550
repeat_region   /note="right ITR"
                /rpt_family="inverted"
gene            5467..6327
                /gene="Amp"
CDS             5467..6327
                /gene="Amp"
                /note="ampicillin resistance"
                /codon_start=1
                /transl_table=11
                /product="beta-lactamase"
                /protein_id="AAV73962.1"
                /db_xref="GI:56119192"
                /translation="MSIGHFRVALIFFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                EIGASLIKHW"
```

```
ORIGIN
```

```
US-10-006-922A-12 (1-225) x AY640630 (1-7147)

Alignment Scores:
Pred. No.:             1.79e-110
Score:                 1186.00              Length:      7147
Percent Similarity:    98.22%               Matches:     219
Best Local Similarity: 97.33%               Conservative: 2
Query Match:           97.69%               Mismatches:  4
                       11                   Indels:      0
                                            Gaps:        0
DB:
```

```
Qy     1 MetArgSerLysGlyAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1168 ATGGGCCTCCCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1109

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1108 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1049

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1048 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 989

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   988 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 929

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   928 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 869

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   868 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 809

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   808 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 749

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   748 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 689

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 42
LOCUS       CQ849511    7616 bp    DNA    linear    PAT   23-AUG-2004
DEFINITION  Sequence 8 from Patent WO2004067751.
ACCESSION   CQ849511
VERSION     CQ849511.1  GI:51507515
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Couillard-Despres,S., Karl,C., Kuhn,H.G. and Aigner,L.
  TITLE     Use of regulatory sequences for specific, transient expression
            in neuronal determined cells
  JOURNAL   Patent: WO 2004067751-A 8 12-AUG-2004;
            Klinikum der Universitaet Regensburg (DE)
FEATURES             Location/Qualifiers
     source          1..7616
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="fusion construct homo sapiens and DsRed 2"
ORIGIN
```

```
US-10-006-922A-12 (1-225) x CQ849511 (1-7616)

Alignment Scores:
Pred. No.:             1.93e-110
Score:                 1186.00              Length:      7616
Percent Similarity:    98.22%               Matches:     219
Best Local Similarity: 97.33%               Conservative: 2
Query Match:           97.69%               Mismatches:  4
                       6                    Indels:      0
                                            Gaps:        0
DB:
```

```
Qy     1 MetArgSerLysGlyAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3606 ATGGGCCTCCCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3665

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3666 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3725

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3726 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3785

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3786 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3845

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3846 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3905

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3906 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3965
```

```
Db   688 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 629

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           ||||||||||||||||||||||||||||||||||:||||||||||:|:|||||||||||||
Db   628 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 569

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   568 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 509

Qy   221 HisHisLeuPheLeu 225
           |||||||||||||||
Db   508 CACCACCTGTTCCTG 494
```

Left column

```
Qy   121 LysValLysPheIleGlyValAsnGlyGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3966 AAGGTGAAGTTCATCGGCGTTGAACTTCCCCTTCCGAGGGCCCGTGATGCAGAAGAAGACC 4025

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4026 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGGGTCCTGAAGGGCGAG 4085

Qy   161 IleHisHisAlaLeuLeuLysAspGlyValHisThrTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4086 ACCCACCACGCCCTGCTGAAGGACGGCGTGCACACCTACCTGGTCGAGTTCAAGTCCATC 4145

Qy   181 TyrMetAlaLeuLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4146 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 4205

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlyAlaLeuGlyTyrArg 220
         ||||||||||||||||||||||||||||||||||||||||||::||||||
Db  4206 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGGAGCCCTGGGCTACCGC 4265

Qy   221 HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db  4266 CACCACCACTTGTTCCTG 4280
```

RESULT 43
CS018252/c

```
LOCUS       CS018252                9468 bp    DNA     linear   PAT 23-FEB-2005
DEFINITION  Sequence 19 from Patent WO2005012534.
ACCESSION   CS018252
VERSION     CS018252.1  GI:6020033
KEYWORDS
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Alphey,L.
  TITLE     Expression systems for insect pest control
  JOURNAL   Patent: WO 200501012534-A 19 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1..9468
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA710"

ORIGIN
```

US-10-006-922A-12 (1-225) x CS018252 (1-9468)

Alignment Scores:
```
Pred. No.:
Score:                   2.52e-110
Score:                   1186.00          Length:        9468
Percent Similarity:      98.22%           Matches:       219
Best Local Similarity:   97.33%           Conservative:  2
Query Match:             97.69%           Mismatches:    4
DB:                                       Indels:        0
                         6                Gaps:          0
```

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||::|||||||||||||||||||||||||||||||||||||||||
Db  8727 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 8668

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||
Db  8667 ACCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 8608

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||
Db  8607 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 8548

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||
Db  8547 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 8488
```

Right column

```
Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  8487 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 8428

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  8427 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 8368

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  8367 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 8308

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  8307 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGGGTCCTGAAGGGCGAG 8248

Qy   161 IleHisHisAlaLeuLeuLysAspGlyValHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  8247 ATCCACCACGCCCTGCTGAAGGACGGCCTGTACCTGGTCGAGTTCAAGTCCATC 8188

Qy   181 TyrMetAlaLeuLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||::|||||||||
Db  8187 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 8128

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlyAlaLeuGlyTyrArg 220
         |||||||||||||||||||||||||||||||||||||::||||||
Db  8127 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGGAGCCCTGGGCTACCGC 8068

Qy   221 HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db  8067 CACCACCACTTGTTCCTG 8053
```

RESULT 44
CS018253/c

```
LOCUS       CS018253               10140 bp    DNA     linear   PAT 23-FEB-2005
DEFINITION  Sequence 20 from Patent WO2005012534.
ACCESSION   CS018253
VERSION     CS018253.1  GI:6020034
KEYWORDS
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Alphey,L.
  TITLE     Expression systems for insect pest control
  JOURNAL   Patent: WO 200501012534-A 20 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1..10140
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA928"

ORIGIN
```

US-10-006-922A-12 (1-225) x CS018253 (1-10140)

Alignment Scores:
```
Pred. No.:
Score:                   2.74e-110
Score:                   1186.00          Length:        10140
Percent Similarity:      98.22%           Matches:       219
Best Local Similarity:   97.33%           Conservative:  2
Query Match:             97.69%           Mismatches:    4
DB:                                       Indels:        0
                         6                Gaps:          0
```

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||::|||||||||||||||||||||||||||||||||||||||||
Db  8727 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 8668

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||
Db  8667 ACCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 8608
```

US-10-006-922A-12 (1-225) x CS018254 (1-10522)

```
DB:
Query Match:        6          Gaps:          0
Best Local Similarity:  97.69%     Indels:        0
Percent Similarity:     97.33%     Mismatches:    4
Score:              1186.00    Conservative:  2
                               Matches:       219
Pred. No.:          2.87e-110  Length:        10522
Alignment Scores:
```

ORIGIN

```
FEATURES         Location/Qualifiers
    source       1..10522
                 /organism="synthetic construct"
                 /mol_type="unassigned DNA"
                 /db_xref="taxon:32630"
                 /note="pLA1124"
SOURCE           synthetic construct
  ORGANISM       other sequences; artificial sequences.
KEYWORDS
VERSION          CS018254.1  GI:60220035
ACCESSION        CS018254
DEFINITION       Sequence 21 from Patent WO2005012534.
LOCUS            CS018254          10522 bp    DNA      linear   PAT 23-FEB-2005
```

RESULT 45

```
Db  8067 CACCACCTGTTCCTG 8053
Qy   221 HisHisLeuPheLeu 225

Db  8127 ATCACCTCCCACACAGGAGGACTACCATCCTGGGACGAGGGCCCGAGGGCCGC 8068
          |||||||||||||||||||:||||||||||||||||||||||||||||||||
Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlnLeuArgThrGluGlyArg 220

Db  8187 TACATGGCCAAGAAGCCCGTGCAGTACGTGGACGCCAAGCTGGAC 8128
          ||||||||||||||||||||:||:|||||||||||||||||||||
Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Db  8247 ACCCACAAGGCCCTGAAGAGCAGCGGCCCCACTACCCTGGAGTTCAAGTCCATC 8188
          |||||||||||||||||||||||||||:||:||||||||||||||||||||||
Qy   161 IleHisLysAlaLeuLysSerSerGlyHisTyrLeuValGlyLeuPheLysSerIle 180

Db  8307 ATGGGCCTGGAGGGCTTCCACCGAGGCCTGTACCCCGCCGAGGGCCGAG 8248
          ||||||||||||||:|||||||||||||||||||||||||||||||||
Qy   141 MetGlyTrpGluLeuAlaSerThrGluArgGlnTyrProArgAspGlyGlyLeuLeuLysGlyGlu 160

Db  8367 AAGGTGAAGTTCATCGGCGTGACAGTTCCTGGGAGCGAGAAGAACC 8308
          ||||||||||||||:|||||||||||:||||||||||||||||||
Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db  8427 GACGGCGGCGTGGCCACCGTGACCCACAGCCTCTCGCTGCTTCATCTAC 8368
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   101 AspGlyGlyValAlaThrValThrHisAspSerSerLeuLeuHisLeuTyr 120

Db  8487 GACTACAAGACGTGCTTCCCCGAGGGCTTCAAGTGGGACACATCCC 8428
          ||||||||||||||||||||||||||||||||||||||||||||||
Qy    81 AspTyrLysThrCysPheProGluGlyPheLysTrpAspThrSerPro 100

Db  8547 CTGTCCCCCAGTTCCAGTACGGCAATCCGTGAAGCACCCCCGACATCCC 8488
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    61 LeuSerProGlnPheGlnTyrGlySerLysSerValThrValLysHisProAlaAspIlePro 80

Db  8607 CACAACACCGTGAAGCTGACCAAGGGGCGGCCCCTTCGCCTGGGACATC 8548
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    41 HisAsnThrValLysLeuThrLysGlyGlyGlyProLeuProPheAlaTrpAspIle 60
```

US-10-006-922A-12 (1-225) x CS018256 (1-10786)

```
DB:
Query Match:        6          Gaps:          0
Best Local Similarity:  97.69%     Indels:        0
Percent Similarity:     97.33%     Mismatches:    4
Score:              1186.00    Conservative:  2
                               Matches:       219
Pred. No.:          2.95e-110  Length:        10786
Alignment Scores:
```

ORIGIN

```
FEATURES         Location/Qualifiers
    source       1..10786
                 /organism="synthetic construct"
                 /mol_type="unassigned DNA"
                 /db_xref="taxon:32630"
                 /note="pLA670"
SOURCE           synthetic construct
  ORGANISM       other sequences; artificial sequences.
KEYWORDS
VERSION          CS018256.1  GI:60220037
ACCESSION        CS018256
DEFINITION       Sequence 23 from Patent WO2005012534.
LOCUS            CS018256          10786 bp    DNA      linear   PAT 23-FEB-2005
```

RESULT 46

CS018256/c

```
Db  5688 CACCACCTGTTCCTG 5702
Qy   221 HisHisLeuPheLeu 225

Db  5628 ATCACCTCCCACACAGGAGGACTACCACCATCCTGGGACGAGGGCCCGAGGGCCGC 5687
          |||||||||||||||||||:||||||||||||||||||||||||||||||||
Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlnLeuArgThrGluGlyArg 220

Db  5568 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 5508
          ||||||||||||||||||||:||:|||||||||||||||||||||
Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Db  5508 ACCCACAAGGCCCTGAAGAGCAGCGGCCACTACCTGGTGGGACTGTTCAAGTCCATC 5567
          |||||||||||||||||||||||||||:||:||||||||||||||||||||||
Qy   161 IleHisLysAlaLeuLysSerSerGlyHisTyrLeuValGlyLeuPheLysSerIle 180

Db  5328 GACGGCGGCGTGGCCACCGTGACCCACGACAGCTCTCCTGCTGCTTCATCTAC 5387
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   101 AspGlyGlyValAlaThrValThrHisAspSerSerLeuLeuHisLeuTyr 120

Db  5208 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCC 5267
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    61 LeuSerProGlnPheGlnTyrGlySerLysSerValThrValLysHisProAlaAspIlePro 80
```

Left column

```
Score:                  1186.00        Matches:        219
Percent Similarity:     98.22%         Conservative:   2
Best Local Similarity:  97.33%         Mismatches:     4
Query Match:            97.69%         Indels:         0
DB:                                    Gaps:           0

US-10-006-922A-12 (1-225) x CS018256 (1-10786)

Qy      1 MetArgSerSerLysArgLysAsnValIleLysGluGluPheMetArgPheLysValAlaArgMetGluGly 20
            |||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db   8727 ATGGCCCTTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGATGGAGGGGC 8668

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8667 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGGGCGGGGGCCCCCTTACGAGGGC 8608

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8607 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGGCCCCTGCCCTTCGCCTGGGACATC 8548

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8547 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 8488

Qy     81 AspThrTyrLysLeuLysPheProGlyPheLysTrpGluArgValMetAsnPheGlu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8487 GACTACAAGAAGCTGTCCTTCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 8428

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8427 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 8368

Qy    121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValLeuLysLysThr 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8367 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGGCCCGTGCTGAAGAAGACC 8308

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db   8307 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 8248

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8247 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 8188

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8187 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 8128

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuTyrGluArgThrGluGlyArg 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8127 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGTACGAGCGCACCGAGGGCCGC 8068

Qy    221 HisHisLeuPheLeu 225
            |||||||||||||||
Db   8067 CACCACCTGTTCCTG 8053
```

```
RESULT 47
LOCUS       CS018251/c                 11251 bp    DNA      linear    PAT    23-FEB-2005
DEFINITION  Sequence 18 from Patent WO2005012534.
ACCESSION   CS018251
VERSION     CS018251.1  GI:60220032
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Alphey,L.
  TITLE     Expression systems for insect pest control
  JOURNAL   Patent: WO 2005012534-A 18 10-FEB-2005;
            Oxitec Limited (GB)
            1. 11251
FEATURES             Location/Qualifiers
     source          1..11251
```

Right column

```
ORIGIN                                  /organism="synthetic construct"
                                        /mol_type="unassigned DNA"
                                        /db_xref="taxon:32630"
                                        /note="pLA656"

US-10-006-922A-12 (1-225) x CS018251 (1-11251)

Alignment Scores:
Pred. No.:              3.11e-110
Score:                  1186.00        Length:         11251
Percent Similarity:     98.22%         Matches:        219
Best Local Similarity:  97.33%         Conservative:   2
Query Match:            97.69%         Mismatches:     4
DB:                                    Indels:         0
                                       Gaps:           0

Qy      1 MetArgSerSerLysArgLysAsnValIleLysGluGluPheMetArgPheLysValAlaArgMetGluGly 20
            |||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db   9191 ATGGCCCTTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGATGGAGGGGC 9132

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9131 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGGGCGGGGGCCCCCTTACGAGGGC 9072

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9071 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGGCCCCTGCCCTTCGCCTGGGACATC 9012

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9011 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 8952

Qy     81 AspThrTyrLysLeuLysPheProGlyPheLysTrpGluArgValMetAsnPheGlu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8951 GACTACAAGAAGCTGTCCTTCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 8892

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8891 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 8832

Qy    121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValLeuLysLysThr 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8831 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGGCCCGTGCTGAAGAAGACC 8772

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db   8771 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 8712

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8711 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 8652

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8651 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 8592

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuTyrGluArgThrGluGlyArg 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8591 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGTACGAGCGCACCGAGGGCCGC 8532

Qy    221 HisHisLeuPheLeu 225
            |||||||||||||||
Db   8531 CACCACCTGTTCCTG 8517
```

```
RESULT 48
LOCUS       CS018250/c                 11570 bp    DNA      linear    PAT    23-FEB-2005
DEFINITION  Sequence 17 from Patent WO2005012534.
ACCESSION   CS018250
VERSION     CS018250.1  GI:60220031
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

RESULT 49

other sequences; artificial sequences.
REFERENCE   1
AUTHORS   Alphey,L.
TITLE   Expression systems for insect pest control
JOURNAL   Patent: WO 2005012534-A 17 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES   Location/Qualifiers
     source   1..11570
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA517"
ORIGIN

US-10-006-922A-12 (1-225) x CS018250 (1-11570)

Alignment Scores:
Pred. No.:
Score: 1186.00
Percent Similarity: 98.22%   Length: 11570
Best Local Similarity: 97.33%   Matches: 219
Query Match: 97.69%   Conservative: 2
DB: Mismatches: 4
                                      Indels: 0
                                      Gaps: 0

Qy   1 MetArgSerLysArgLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||::|||||||||||||||||||||||||
Db 8886 ATGGGCCTTCCTCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 8827

Qy   21 ThrValAsnGlyHisGlyLeuPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8826 ACCGTGAACGGCCACGGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 8767

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8766 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 8707

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8706 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 8647

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8646 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 8587

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8586 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 8527

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8526 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTGATGCAGAAGAAGACC 8467

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8466 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 8407

Qy 161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8406 ATCCACCACGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 8347

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db 8346 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 8287

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8286 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 8227

Qy 221 HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db 8226 CACCACCTGTTCCTG 8212

LOCUS   CS018255   11867 bp   DNA   linear   PAT 23-FEB-2005
DEFINITION Sequence 22 from Patent WO2005012534.
ACCESSION   CS018255
VERSION   CS018255.1 GI:60220036
KEYWORDS   .
SOURCE   synthetic construct
ORGANISM   synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
AUTHORS   Alphey,L.
TITLE   Expression systems for insect pest control
JOURNAL   Patent: WO 2005012534-A 22 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES   Location/Qualifiers
     source   1..11867
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA1188"
ORIGIN

US-10-006-922A-12 (1-225) x CS018255 (1-11867)

Alignment Scores:
Pred. No.:
Score: 1186.00
Percent Similarity: 98.22%   Length: 11867
Best Local Similarity: 97.33%   Matches: 219
Query Match: 97.69%   Conservative: 2
DB: Mismatches: 4
                                      Indels: 0
                                      Gaps: 0

Qy   1 MetArgSerLysArgLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||::|||||||||||||||||||||||||
Db 6373 ATGGGCCTTCCTCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 6432

Qy   21 ThrValAsnGlyHisGlyLeuPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6433 ACCGTGAACGGCCACGGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 6492

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6493 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 6552

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6553 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 6612

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6613 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 6672

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6673 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 6732

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6733 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTGATGCAGAAGAAGACC 6792

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6793 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 6852

Qy 161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6853 ATCCACCACGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 6912

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db 6913 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 6972

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220

RESULT 50

```
LOCUS       CS018249              11920 bp    DNA     linear   PAT 23-FEB-2005
CS018249
DEFINITION  Sequence 16 from Patent WO2005012534.
ACCESSION   CS018249
VERSION     CS018249.1  GI:60220030
KEYWORDS    
SOURCE      Oxitec Limited (GB)
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  11920
  AUTHORS   Alphey,L.
  TITLE     Expression systems for insect pest control
  JOURNAL   Patent: WO 2005012534-A 16 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1..11920
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA513"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                  3.34e-110
                        1186.00
Percent Similarity:     98.22%
Best Local Similarity:  97.33%
Query Match:            97.69%
DB:                     6

Length:        11920
Matches:       219
Conservative:  
Mismatches:    2
Indels:        4
Gaps:          0
```

US-10-006-922A-12 (1-225) x CS018249 (1-11920)

```
Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8013 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTTACGAGGGC 8072

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   8073 CACAACACCGTGAAGCGTCATCACCCGAGTTCATGCCGCTTCAAGGTTCCGCATGGAGGGC 8012

Qy     41 HisAsnThrValLysArgValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8073 CACAACACCGTGAAGCGTCATCACCCGAGTTCATGCCGCTTCAAGGTTCCGCATGGAGGGC 8132

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8133 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCAACATCCCC 8192

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8193 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 8252

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8253 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 8312

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8313 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 8372

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8373 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 8432

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 51

```
LOCUS       CS018257              14720 bp    DNA     linear   PAT 23-FEB-2005
CS018257/c
DEFINITION  Sequence 24 from Patent WO2005012534.
ACCESSION   CS018257
VERSION     CS018257.1  GI:60220038
KEYWORDS    
SOURCE      Oxitec Limited (GB)
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  14720
  AUTHORS   Alphey,L.
  TITLE     Expression systems for insect pest control
  JOURNAL   Patent: WO 2005012534-A 24 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1..14720
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA1038"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                  4.32e-110
                        1186.00
Percent Similarity:     98.22%
Best Local Similarity:  97.33%
Query Match:            97.69%
DB:                     6

Length:        14720
Matches:       219
Conservative:  
Mismatches:    2
Indels:        4
Gaps:          0
```

US-10-006-922A-12 (1-225) x CS018257 (1-14720)

```
Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7382 ATGGCCCTCCTCCGAGAACGTCATCACCCGAGTTCATGCCGCTTCAAGGTGCCGCATGGAGGGC 7323

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   7322 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTTACGAGGGC 7263

Qy     41 HisAsnThrValLysArgValThrLysGlyGlyProLeuProPheAlaTrpAspIlePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7262 CACAACACCGTGAAGCGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 7203

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7202 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCAACATCCCC 7143

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7142 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 7083

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7082 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 7023

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   6973 ATCACCTCCCACACGAGGACTACACCATCGGCTGGGAGGCAGTAACGACGGAGGACCGGC 7032
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    221 HisHisLeuPheLeu 225
          |||||||||
Db   7033 CACCACCTGTTCCTG 7047
```

```
Db   8433 ACCCACAAAGGCCCTGAAGCTGAAGGACGGCGGCCGGCCACTACCTGGTGGAGTTCAAGTCCATC 8492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAspSerLysLeuAsp 200
          ||||||||||||||  |||||||||||||||||||||| :||||||||||||||||||||||
Db   8493 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGACAGCAAGCTGGAC 8552

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8553 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 8612

Qy    221 HisHisLeuPheLeu 225
          |||||||||
Db   8613 CACCACCTGTTCCTG 8627
```

```
Db   7022  AAGGTGAAGTTCATCCGAGTCGAGATCGAGAAGCTCCGGACGGCCCGGGATGCTCGAGAAGAAGACC  6963
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6962  ATGGGCTGGGAGGCCTTCACCGAGCCCGTGCTCTAAAGGGCGAG  6903
Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAlaAspGlyValLeuLysGlyGlu  160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6902  ACCGACAAGGCCCTGAAGCTGCAGAGAAGACGGGGGGGCCCACTACGTTCAGTCCATC  6843
Qy    161  IleHisLysAlaLeuLysLeuLysGlyAspGlyValHisTyrLeuValGluPheLysSerIle  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6842  TACACGGGCCAAGAAGCCCCGTTGCCAGCTGGGGGGGCC  6783
Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAlaAspSerLysLeuAsp  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6782  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGGTGACCAGAACGGAAGGGCCGGC  6723
Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluValThrArgThrGluGlyAsnArg  220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6722  CACCACCTGTTCCTG  6708
Qy    221  HisHisLeuPheLeu  225
```

RESULT 52
CQ981073/c

```
LOCUS       CQ981073               16157 bp    DNA     linear   PAT 25-JAN-2005
DEFINITION  Sequence 3 from Patent WO2005003364.
ACCESSION   CQ981073
VERSION     CQ981073.1  GI:58190337
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (residues 1 to 16157)
  AUTHORS   Alphey,L.
  TITLE     Stable integrands
  JOURNAL   Patent: WO 2005003364-A 3  13-JAN-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1..16157
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="predicted sequence of pLA1025 construct"
ORIGIN
```

```
Alignment Scores:

Pred. No.:                  1
Score:                4.83e-110
Percent Similarity:      1186.00
Best Local Similarity:    98.22%
Query Match:              97.33%
DB:                       97.69%
```

US-10-006-922A-12 (1-225) x CQ981073 (1-16157)

```
                          Length:      16157
                          Matches:       219
                          Conservative:    2
                          Mismatches:      4
                          Indels:          0
                          Gaps:            0
```

```
Qy      1  MetArgSerSerLysSerAsnValIleLeuLysGluMetArgPheLysValArgMetGluGly  20
           |||||||||||||||||||||||||||||||||||||||||||||::|||||||||||||||||
Db  14524  ATGGGCTCCTCCAAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  14465
Qy     21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14464  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  14405
Qy     41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14404  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  14345
Qy     61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14344  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  14285
Qy     81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  14284  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG  14225
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
           ||||||||||||||||||||||||||||||||::||||||||||||||||||||||||||
Db  14224  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  14165
Qy    121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14164  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC  14105
Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAlaAspGlyValLeuLysGlyGlu  160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14104  ATGGGCTGGGAGGCCTTCACCGAGCGCCTGTACCCCCGCGCCGACGGCGTGCTGAAGGGC  14045
Qy    161  IleHisLysAlaLeuLysLeuLysGlyAspGlyValHisTyrLeuValGluPheLysSerIle  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14044  ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  13985
Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAlaAspSerLysLeuAsp  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db  13984  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGCCGACAGCAAGCTGGAC  13925
Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluValThrArgThrGluGlyAsnArg  220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13924  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGGTGACCAGAACCGAGGGCAACCGC  13865
Qy    221  HisHisLeuPheLeu  225
           |||||||||||||||
Db  13864  CACCACCTGTTCCTG  13850
```

RESULT 53
CQ981074/c

```
LOCUS       CQ981074               16157 bp    DNA     linear   PAT 25-JAN-2005
DEFINITION  Sequence 4 from Patent WO2005003364.
ACCESSION   CQ981074
VERSION     CQ981074.1  GI:58190338
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (residues 1 to 16157)
  AUTHORS   Alphey,L.
  TITLE     Stable integrands
  JOURNAL   Patent: WO 2005003364-A 4  13-JAN-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1..16157
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="predicted sequence of pLA1125 construct"
ORIGIN
```

```
Alignment Scores:

Pred. No.:                  1
Score:                4.83e-110
Percent Similarity:      1186.00
Percent Similarity:      98.22%
Best Local Similarity:    97.33%
Query Match:              97.69%
```

US-10-006-922A-12 (1-225) x CQ981074 (1-16157)

```
                          Length:      16157
                          Matches:       219
                          Conservative:    2
                          Mismatches:      4
                          Indels:          0
                          Gaps:            0
```

```
Qy      1  MetArgSerSerLysSerAsnValIleLeuLysGluMetArgPheLysValArgMetGluGly  20
           |||||||||||||||||||||||||||||||||||||||||||||::|||||||||||||||||
Db  14524  ATGGGCTCCTCCAAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  14465
Qy     21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14464  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  14405
Qy     41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14404  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  14345
```

```
US-10-006-922A-12 (1-225) x AX686888 (1-898)

DB:
Query Match:            97.53%
Best Local Similarity:  97.33%
Percent Similarity:     98.22%
Score:                  1184.00
Pred. No.:              2.28e-111
Alignment Scores:

ORIGIN

FEATURES             Location/Qualifiers
     source          1..898
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"

SOURCE
     ORGANISM  Discosoma sp.
               Discosoma sp.
               Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
               Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE    1
     AUTHORS   Lukyanov,S.A., Fradkov,A.F., Labas,Y.A., Matz,M.V. and Terskikh,A.
     TITLE     Anthozoa derived chromo/fluoroproteins and methods for using the
               same
     JOURNAL   Patent: WO 0127150-A 11 19-APR-2001;
               Clontech Laboratories Inc. (US)
KEYWORDS
VERSION      AX686888.1  GI:29409468
ACCESSION    AX686888
DEFINITION   Sequence 11 from Patent WO0127150.
LOCUS        AX686888          898 bp    DNA    linear    PAT 31-MAR-2003

RESULT 54

Db    13864 CACCACCTGTTCCTG 13850
            |||||||||||||||
Qy      221 HisHisLeuPheLeu 225

Db    13924 ATCACCTCCCACACGAGGACTACCATCGTGCGACCGAGGGCGCC 13865
            |||||||||||||||||||||||||||||||||||||||||||
Qy      201 IleThrSerHisAsnGluAspThrIleValGluGlnArgThrValGluGlyAla 220

Db    13984 TACATGGCCAAGAGAGCCCGTGAGCTGCCCGGCTACTACTACGTTGACAGCCAAGCTGGAC 13925
            ||||||||||||||||||||||||||||| ::::||||||||||||||||||||||||
Qy      181 TyrMetAlaLysArgAlaProGluLeuProGlyTyrTyrTyrValAspSerGlnLeuAsp 200

Db    14044 ACCCACAACAGGCCCTGAAGCTGAAGGACGGGGGCCACTACCTGGTGGAGTTCAAGTCCATC 13985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db    14104 ATGGGCTGGGAGCCCTCCACCGAGCGCCTGTACCCGGCGGATGGCCTGAAAGGCGAG 14045
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      141 MetGlyTrpGluProSerThrGluArgLeuTyrProAlaAspGlyLeuLysGlyGlu 160

Db    14164 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 14105
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db    14224 GACGGCGGCGGCGTGGCCACGGTGACCCAGGACACCAGCCTGCTGCTTCACTAC 14165
            |||||||||||||||||||||||||||||||||||||||||||||
Qy      101 AspGlyGlyGlyValAlaThrValThrGlnAspThrSerLeuLeuLeuHisTyr 120

Db    14284 GACTACAAGAAGAGCTGTCCTTCCCCAGAGGGGTTCAAGTGGGAGCGCGTGATGAACTTCGAG 14225
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       81 AspTyrLysLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

Db    14344 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCGGACATCCCC 14285
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db    14404 CACGACACCCGTGAAGCTGAAGGTGAACCAAGGGCGCCCCCTTCGCCCTGGGACATC 14345
            ||||||||||||||||||||||||||||||||||||||||
Qy
```

```
FEATURES             Location/Qualifiers
     source          1..3441
                     /organism="Dual fluorescent protein cloning vector pGRFP"
                     /mol_type="other DNA"
JOURNAL   Submitted (01-FEB-2005) Laboratory of Ger van den Engh, Institute
          for Systems Biology, 1441 N 34th St, Seattle, WA 98103-8904, USA
TITLE     Direct Submission
AUTHORS   Choe,J., Guo,H.H. and van den Engh,G.
REFERENCE    2  (bases 1 to 3441)
PUBMED    15767274
JOURNAL   (er) Nucleic Acids Res. 33 (5), E49 (2005)
          characterization and selection by cell sorting
          A dual-fluorescence reporter system for high-throughput clone
TITLE     Choe,J. Guo,H.H. and van den Engh,G.
AUTHORS   1  (bases 1 to 3441)
REFERENCE    1
          other sequences; artificial sequences; vectors.
          Dual fluorescent protein cloning vector pGRFP
ORGANISM  Dual fluorescent protein cloning vector pGRFP
SOURCE
KEYWORDS
VERSION   AY916793.1  GI:60101687
ACCESSION AY916793
DEFINITION Dual fluorescent protein cloning vector pGRFP, complete sequence.
LOCUS     AY916793          3441 bp    DNA    circular SYN 30-MAR-2005

RESULT 55

Db      753 CACCCATCTGTTCCTT 767
            |||||||||||||||
Qy      221 HisHisLeuPheLeu 225

Db      573 ATTCATAAGGCTCTGAAGCTGAAGACGGTGGTCCATTACCTAGTTGAATTCAAAAGTATT 632
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db      513 ATGGGCTGGGAACCCAGCACGGAACGGCTTTATCCTGCAGACGGTTTGAAAGGAGAG 572
            |||||||||||||||||||||||||||||||||||||||||||||
Qy      141 MetGlyTrpGluProSerThrGluArgLeuTyrProAlaAspGlyLeuLysGlyGlu 160

Db      453 AAGTCAAGTTCAATTGGCGTTAACTTTCCTTCGGATGGACCTGTTATGCAAAAGAAGACA 512
            ||| :::  ||||||||||||||||||||||||||||||||||||||||||||||||
Qy      121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db      393 GACGGTGGCGGTCGTTACTGTAACCCAGGATTCCAGTTGCAGGATGGGCTCAACTCTAC 452
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      101 AspGlyGlyGlyValAlaThrValThrGlnAspThrSerLeuLeuLeuHisTyr 120

Db      333 GACTATAAAAAGGGTCTTTCCTTCCCAGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 392
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       81 AspTyrLysLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

Db      273 TTGTCACCACAATTTCAGTATGGAAGCAAGGTTTATGTCAAGCACCCTGCCGACATACCA 332
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db      213 CACAATACCGTAAAGCTTAAGGTAACCAAAGGGGGACCTTGCCATTTGCTTGGGGGATATT 272
            ||||||||||||||||||||||||||||||||||||||||||||||
Qy       41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProLeuAlaTrpAspIle 60

Db      153 ACGGTCAATGGCCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGGAGAGGC 212
            ||||||||||||||||||||||||||||||||||||||||||||
Qy       21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40

Db       93 ATGAGGTCTTCAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTGCAAGGGGAGGA 152
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
```

gene   /db_xref="taxon:314306"
       /gene="GRFP"
       140. .1630

CDS   /gene="GRFP"
       140. .1630
       /note="fusion protein containing GFPmut3.1 and DsRed-T3
       with an intervening linker region"
       /codon_start=1
       /transl_table=11
       /product="GFP-DsRed fusion protein"
       /protein_id="AAX13969.1"
       /db_xref="GI:60101689"
       /translation="MRKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTL
       KFICTTGKLPVPWPTLVTTFGYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKD
       DGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNG
       IKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHM
       VLLEFVTAAGITHGMDELYKSGSGSGCKKTTASSAAADISVDGHSCFLGSSGSGMASSE
       DVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSP
       QFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYK
       VKFIGVNFPSDGPVMQKKTMGWEPSTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKS
       IYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTEGRHHLFL"

misc_feature   /gene="GRFP"
               140. .853
               /note="faster folding mutant of discosoma red fluorescent
               protein; derived from pdsRed-T3; silent mutations have
               been introduced to change restriction sites; Region:
               discosoma red fluorescent protein T3 mutant."

misc_feature   1778. .2638
               /function="ampicillin resistance"
               /codon_start=1
               /transl_table=11
               /product="beta-lactamase"
               /protein_id="AAX13968.1"
               /db_xref="GI:60101688"
               /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
               IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
               YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
               DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
               LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
               EIGASLIKHW"

misc_feature   854. .952
               /gene="GRFP"
               /note="cycle 3 mutant of GFP; derived initially from
               pGFPmut3.1; Region: green fluorescent protein mutant 3.1"

misc_feature   953. .1627
               /gene="GRFP"
               /note="contains G,S residues for flexibility on each side,
               and NotI, EcoRV, SalI cloning sites; Region: linker"

ORIGIN

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

US-10-006-922A-12 (1-225) x AY916793 (1-3441)

Score: 3.16e-109   Length: 3441
Percent Similarity: 97.78%   Matches: 217
                      96.44%   Conservative: 3
Best Local Similarity: 96.44%   Mismatches: 5
Query Match: 96.38%   Indels: 0
DB: 11   Gaps: 0

Qy   1 MetArgSerLysGluGluLeuPheThrGlyValValProIleLeuLysGluMetArgPheLysValArgMetGluGluGly   20
Db   953 ATGGGCTCTCCGGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGAGGGC 1012

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db 1013 TCCGTGAACGGCCACGAGTTCGAGAATCGAGGGCGAGGGCGCCCCCTACGAGGGC 1072

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db 1073 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1132

US-10-006-922a-12.p2nminscr.rge   Page 38

RESULT 56

LOCUS   AJ851285   5311 bp   DNA   circular SYN 12-NOV-2004
DEFINITION Cloning vector pRU1105 dsRedT3 gene for Ds Red fluorescent protein
            T3.
ACCESSION   AJ851285
VERSION   AJ851285.1 GI:55724876
KEYWORDS   Ds Red fluorescent protein T3; dsRedT3 gene.
SOURCE   Cloning vector pRU1105
  ORGANISM Cloning vector pRU1105
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE   High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 5311)
  AUTHORS   Poole,P.S.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES   Location/Qualifiers
     source   1..5311
                     /organism="Cloning vector pRU1105"
                     /mol_type="other DNA"
                     /db_xref="taxon:299180"
                     /lab_host="Escherichia coli"
     gene   326..1003
                     /gene="dsRedT3"
     CDS   326..1003
                     /gene="dsRedT3"
                     /codon_start=1
                     /transl_table=11
                     /product="Ds Red fluorescent protein T3"
                     /protein_id="CAH64890.1"
                     /db_xref="GI:55724877"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db 1133 CTGTCCCCGCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1192

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 1193 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1252

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 1253 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1312

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 1313 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 1372

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 1373 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1432

Qy   161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 1433 ATCCACCACGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1492

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 1493 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1552

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 1553 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCAGC 1612

Qy   221 HisHisHisLeuPheLeu 225
Db 1613 CACCACCATCTGTTCCTT 1627

```
ORIGIN
        GVVTVTQDDSLQDGCFIYVKFIGVNPPSDGPVMQKKTMGWEPSTERLYPRDGVLKGE
        IHKALKLKDGGHYLVEFRKSIYMAKKPVQLPGYYVVDSKLDITSHNEDYTIVEQYERTE
        GRHHLFL"

US-10-006-922a-12 (1-225) x AJ851285 (1-5311)

Alignment Scores:
Pred. No.:
Score:                  5.36e-109   Length:        5311
                        1170.00     Matches:       217
Percent Similarity:     97.78%      Conservative:  3
Best Local Similarity:  96.44%      Mismatches:    5
Query Match:            96.38%      Indels:        0
DB:                     11          Gaps:          0

Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        |||..|||||||:||::||||||||||||||||||||||||||||||||||||
Db  326 ATGGCCTTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCCCATGGAGGGC 385

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        :::||.||||||||||||||||||||||||||||||||||||||||||||||||||
Db  386 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 445

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||
Db  446 ACCCAGACACCCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 505

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  506 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 565

Qy   81 AspGlyTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  566 GACGGCTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 625

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  626 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 685

Qy  121 LysValLysPheIleGlyValAsnProProSerAspGlyProValMetGlnLysLysThr 140
Db  686 AAGGTGAAGTTCATCGGCGTGAACCCTCCCTCGGACGGCCCCGTAATGCAGAAGAAGACT 745

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  746 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTCCTGAAGGGCGAG 805

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  806 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 865

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  866 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 925

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  926 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 985

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||
Db  986 CACCACCTGTTCCTG 1000

RESULT 57
LOCUS       AJ851289          13079 bp    DNA    circular SYN 12-NOV-2004
DEFINITION  Cloning vector pRU1075 guA gene for beta-glucuronidase and dsRedT3
            gene for Ds Red fluorescent protein T3.
ACCESSION   AJ851289
VERSION     AJ851289.1  GI:55724885
KEYWORDS    beta-glucuronidase; Ds Red fluorescent protein T3; dsRedT3 gene;
            guA gene.
SOURCE      Cloning vector pRU1075
```

```
  ORGANISM  Cloning vector pRU1075
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE     High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13079)
  AUTHORS   Poole,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..13079
                     /organism="Cloning vector pRU1075"
                     /mol_type="other DNA"
                     /db_xref="taxon:299184"
                     /lab_host="Escherichia coli"
     gene            186..863
                     /gene="dsRedT3"
     CDS             186..863
                     /gene="dsRedT3"
                     /codon_start=1
                     /transl_table=11
                     /product="Ds Red fluorescent protein T3"
                     /protein_id="CAH64917.1"
                     /db_xref="GI:55724886"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVVTVTQDDSLQDGCFIYKVKFIGVNPPSDGPVMQKKTMGWEPSTERLYPRDGVLKGE
                     IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERTE
                     GRHHLFL"
     gene            885..2696
                     /gene="guaA"
     CDS             885..2696
                     /gene="guaA"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucuronidase"
                     /protein_id="CAH64918.1"
                     /db_xref="GI:55724887"
                     /translation="MVRPVEPTPREIKKLDGLWAFSLDRENCGIDQRWWESALQESRA
                     IAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
                     EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTIPPGMVITDENGKKKQSY
                     FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
                     ADQQVVATGQGTSGTLQVVNPHLWQPGEGYLYELCVTAKSQTECDIYPLRVGIRSVAV
                     KGEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
                     EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAI
                     KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
                     AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVD
                     TLAGLHSMYTDMWSEEYQCAWLDMYHRVFDRVSAAVGEQVWNFADFATSQGILRVGGN
                     KKGIFTRDRKPKSAAFLLQKRWTGMNFGEKPQQGGKQ"

US-10-006-922a-12 (1-225) x AJ851289 (1-13079)

Alignment Scores:
Pred. No.:
Score:                  1.61e-108   Length:        13079
                        1170.00     Matches:       217
Percent Similarity:     97.78%      Conservative:  3
Best Local Similarity:  96.44%      Mismatches:    5
Query Match:            96.38%      Indels:        0
DB:                     11          Gaps:          0

Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        |||..|||||||:||::||||||||||||||||||||||||||||||||||||
Db  186 ATGGCCTTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCCCATGGAGGGC 245

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        :::||.||||||||||||||||||||||||||||||||||||||||||||||||||
Db  246 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 305

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||
Db
```

```
Db   306  ACCCAGACACGCCAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  365
Qy    61  LeuSerProGlnPheGlnTyrGlySerLeuValTyrValValHisProAlaAspIlePro   80
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db   366  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGGTGCACCCCGCCGACATCCCC  425
Qy    81  AspThrTyrLysLeuSerPhePheLysGlyPheGluGlyValMetAsnGlu           100
          |||||  |||||||||||||  |||||||||||||||||||||||
Db   426  GACTACAAGAGCTGTCCTTCCCCAAGGGCTTCGAGGGCGTGATGAACTTCGAG          485
Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   486  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  545
Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   546  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT  605
Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
          ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db   606  ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  665
Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   666  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  725
Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   726  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  785
Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   786  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  845
Qy   221  HisHisLeuPheLeu  225
          |||||||||||||||
Db   846  CACCACCTGTTCCTG  860
```

```
RESULT 58
AJ851286

LOCUS       AJ851286                5311 bp    DNA     circular SYN 12-NOV-2004
DEFINITION  Cloning vector pRU1106 dsRedT4 gene for Ds Red fluorescent protein
            T4.
ACCESSION   AJ851286
VERSION     AJ851286.1  GI:5572487
KEYWORDS    Ds Red fluorescent protein T4; dsRedT4 gene.
SOURCE      Cloning vector pRU1106
  ORGANISM  Cloning vector pRU1106
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 5311)
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE     High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5311)
  AUTHORS   Poole,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..5311
                     /organism="Cloning vector pRU1106"
                     /mol_type="other DNA"
                     /db_xref="taxon:299181"
                     /lab_host="Escherichia coli"
     gene            326..1003
                     /gene="dsRedT4"
     CDS             326..1003
                     /gene="dsRedT4"
                     /codon_start=1
                     /transl_table=11
                     /product="Ds Red fluorescent protein T4"
                     /protein_id="CAH64891.1"
                     /db_xref="GI:5572487879"
```

```
ORIGIN
        /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
        KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
        GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEPSTERLYPRDGVLKGE
        IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAE
        GRHHLFL"
```

```
US-10-006-922A-12 (1-225) x AJ851286 (1-5311)

Alignment Scores:
Pred. No.:            1.74e-108   Length:       5311
Score:                1165.00     Matches:      216
Percent Similarity:   97.33%      Conservative: 3
Best Local Similarity: 96.00%     Mismatches:   6
Query Match:          95.96%      Indels:       0
DB:                   11          Gaps:         0
```

```
Qy     1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
          |||
Db   326  ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  385
Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly    40
          |::||||||||||||||||||||||||
Db   386  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  445
Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          |||
Db   446  ACCCAGACACGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  505
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          |||
Db   506  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  565
Qy    81  AspTyrLysLysLeuSerPhePheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
          ||||||||
Db   566  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  625
Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          |||
Db   626  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  685
Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
          |||
Db   686  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT  745
Qy   141  MetGlyTrpGluProSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
          |||
Db   746  ATGGGCTGGGAGCCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  805
Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
          |||
Db   806  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  865
Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
          |||
Db   866  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  925
Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
          |||
Db   926  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  985
Qy   221  HisHisLeuPheLeu  225
          |||
Db   986  CACCACCTGTTCCTG  1000
```

```
RESULT 59
AY342347

LOCUS       AY342347               10141 bp    DNA     linear   SYN 21-OCT-2003
DEFINITION  Red H-Pelican DsRed.T4 transformation vector, complete sequence.
ACCESSION   AY342347
VERSION     AY342347.1  GI:33358309
KEYWORDS    .
SOURCE      Red H-Pelican DsRed.T4 transformation vector
```

Left column:

ORGANISM	Red H-Pelican DsRed.T4 transformation vector other sequences; artificial sequences; vectors.
REFERENCE	1 (bases 1 to 10141)
AUTHORS	Barolo,S. and Posakony,J.W.
TITLE	Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP [DsRed.T4] and Nuclear RFP in Insulated Vectors
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 10141)
AUTHORS	Barolo,S. and Posakony,J.W.
TITLE	Direct Submission
JOURNAL	Submitted (14-JUL-2003) Division of Biology/CDB, UCSD, 4121 Bonner Hall, MC 0349, La Jolla, CA 92093-0349, USA
FEATURES	Location/Qualifiers
source	1..10141
	/organism="Red H-Pelican DsRed.T4 transformation vector"
	/mol_type="other DNA"
	/db_xref="taxon:240544"
ORIGIN	

US-10-006-922A-12 (1-225) x AY342347 (1-10141)

Alignment Scores:
Pred. No.:
Score: 3.82e-108   Length: 10141
Percent Similarity: 97.33%   Matches: 216
Best Local Similarity: 96.00%   Conservative: 3
Query Match: 95.96%   Mismatches: 6
    Indels: 0
DB: 11   Gaps: 0

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGlyGly 20
        ||| :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  285 ATGGCCTCTCCAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 344

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
        ::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  345 TCCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 404

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  405 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 464

Qy   61 LeuSerProGlnPheGlnTyrGlySerIleTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  465 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 524

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  525 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 584

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  585 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 644

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  645 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 704

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  705 ATGGGCTGGGAGGCCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 764

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  765 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 824

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  825 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 884

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  885 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 944
```

Right column:

```
Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  945 CACCACCTGTTCCTG 959
```

RESULT 60
```
AY342348    AY342348    10276 bp    DNA    linear    SYN 21-OCT-2003
LOCUS    AY342348    Red H-Stinger DsRed.T4-NLS transformation vector, complete
DEFINITION    sequence.
ACCESSION    AY342348
VERSION    AY342348.1    GI:33358310
KEYWORDS    
SOURCE    Red H-Stinger DsRed.T4-NLS transformation vector
ORGANISM    Red H-Stinger DsRed.T4-NLS transformation vector
        other sequences; artificial sequences; vectors.
REFERENCE    1 (bases 1 to 10276)
AUTHORS    Barolo,S. and Posakony,J.W.
TITLE    Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP
        [DsRed.T4] and Nuclear RFP in Insulated Vectors
JOURNAL    Unpublished
REFERENCE    2 (bases 1 to 10276)
AUTHORS    Barolo,S. and Posakony,J.W.
TITLE    Direct Submission
JOURNAL    Submitted (14-JUL-2003) Division of Biology/CDB, UCSD, 4121 Bonner
        Hall, MC 0349, La Jolla, CA 92093-0349, USA
FEATURES    Location/Qualifiers
source    1..10276
        /organism="Red H-Stinger DsRed.T4-NLS transformation
        vector"
        /mol_type="other DNA"
        /db_xref="taxon:240545"
ORIGIN
```

US-10-006-922A-12 (1-225) x AY342348 (1-10276)

Alignment Scores:
Pred. No.:
Score: 3.88e-108   Length: 10276
Percent Similarity: 97.33%   Matches: 216
Best Local Similarity: 96.00%   Conservative: 3
Query Match: 95.96%   Mismatches: 6
    Indels: 0
DB: 11   Gaps: 0

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGlyGly 20
        ||| :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  288 ATGGCCTCTCCAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 347

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
        ::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  348 TCCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 407

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  408 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 467

Qy   61 LeuSerProGlnPheGlnTyrGlySerIleTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  468 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 527

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  528 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 587

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  588 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 647

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  648 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 707

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||
Db  648 ...
```

RESULT 61

```
LOCUS       AY490568               10481 bp    DNA     linear   SYN 24-DEC-2003
DEFINITION  UAS-Red Stinger DsRed.T4-NLS Drosophila UAS-RF transformation
            vector, complete sequence.
ACCESSION   AY490568
VERSION     AY490568.1  GI:40218087
KEYWORDS    .
SOURCE      UAS-Red Stinger DsRed.T4-NLS Drosophila UAS-RF transformation
            vector
  ORGANISM  UAS-Red Stinger DsRed.T4-NLS Drosophila UAS-RF transformation
            vector
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 10481)
  AUTHORS   Barolo,S. and Posakony,J.W.
  TITLE     Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP
            [DsRed.T4] and Nuclear RFP in Insulated Vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10481)
  AUTHORS   Barolo,S. and Posakony,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2003) Cell & Developmental Biology, University of
            Michigan Medical School, 5732 Med Sci 2, Ann Arbor, MI 48109, USA
FEATURES             Location/Qualifiers
     source          1..10481
                     /organism="UAS-Red Stinger DsRed.T4-NLS Drosophila UAS-RF
                     transformation vector"
                     /mol_type="other DNA"
                     /db_xref="taxon:258595"
```

```
US-10-006-922A-12 (1-225) x AY490568 (1-10481)

Alignment Scores:
Pred. No.:
Score:                     1165.00
Percent Similarity:          97.33%       Length:        10481
Best Local Similarity:       96.00%       Matches:         216
Query Match:                 95.96%       Conservative:      3
DB:                          11           Mismatches:        6
                                          Indels:            0
                                          Gaps:              0

ORIGIN

Qy    1 MetArgSerLysAsnValIleLeuLysGlyLeuAsnValIleLeuGluMetGluGly 20
        ||| |||||||||||::::|||| |||::||| |||:::| |:
Db  494 ATGGCCTCCTTCGGAGGACGTCATCAAGGGAGTTCATCGGCTTCAAGGAGTGCGGGG 553

Qy   21 ThrValAsnGlyHisGlyPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  554 TCCGTGAACGGCCACCAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 613

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  614 ACCCAGACACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATC 673

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  674 ...

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  768 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 827

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  828 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 887

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  888 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACGGAGGGCCGC 947

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  948 CACCACCTGTTCCTG 962
```

(Db 708 ATGGGGCTGGGAGCCCTCCACCGAGGCCCTGTACCCCGGGACCGGGTGCTGAAGGGGCGAG 767)

RESULT 62

```
LOCUS       AJ851290               13079 bp    DNA     circular SYN 12-NOV-2004
DEFINITION  Cloning vector pRU1076 gusA gene for beta-glucuronidase and dsRed
            gene for Ds Red fluorescent protein T4.
ACCESSION   AJ851290
VERSION     AJ851290.1  GI:55724888
KEYWORDS    beta-glucuronidase; Ds Red fluorescent protein T4; dsRed gene;
            gusA gene.
SOURCE      Cloning vector pRU1076
  ORGANISM  Cloning vector pRU1076
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE     High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13079)
  AUTHORS   Poole,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..13079
                     /organism="Cloning vector pRU1076"
                     /mol_type="other DNA"
                     /db_xref="taxon:299185"
                     /lab_host="Escherichia coli"
     gene            186..863
                     /gene="dsRedT4"
     CDS             186..863
                     /gene="dsRedT4"
                     /codon_start=1
                     /transl_table=11
                     /product="Ds Red fluorescent protein T4"
                     /protein_id="CAH64919.1"
                     /db_xref="GI:55724889"
                     /db_xref="GOA:Q5VYKF"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGE
```

```
Db  674 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCGGCCGACATCCCC 733

Qy   81 AspThrTyrLysSerSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  734 GACTACAAGAAGAGCAGCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 793

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  794 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 853

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  854 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 913

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  914 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 973

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  974 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1033

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1034 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1093

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1094 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACGGAGGGCCGC 1153

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 1154 CACCACCTGTTCCTG 1168
```

```
IHKALKLKDGGHYLVEFKSIYMAKKPVQLPGYYYVDSKLDITSHNEDYTIVEQYERAE
GRHHLFL"
```

```
     gene            885..2696
                     /gene="guaA"
     CDS             885..2696
                     /gene="guaA"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucuronidase"
                     /protein_id="CAH64920.1"
                     /db_xref="GI:55724890"
                     /translation="MVRPVETPTREIKLDGLWAFSLDRENCGIDQRWWESALQESRA
     IAVPGSFNDQFADAIRNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
     EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVANNELNWQTIPPGMVIDENGKKKQSY
     FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
     ADQQVATCQGTSGTLQVVNPHLWQPGEGYLYELCVTAKSQTECDIYPLRVGIRSVAV
     KGEQFLINHKPFYFTGFGRHEDADILRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
     EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQRAHLQAI
     KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
     AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVD
     TLAGLHSMYTDMWSEEYQCAWLDMYHRVFDRVSAVVGEQVWNFADFATSQGILRVGGN
     KKGIFTRDRKPKSAAFLLQKRWTGMNFGEKPQQGGKQ"
```

US-10-006-922A-12 (1-225) x AJ851290 (1-13079)

ORIGIN

Alignment Scores:
Pred. No.:
Score: 5.21e-108
Percent Similarity: 1165.00
Best Local Similarity: 97.33%
Query Match: 96.00%
                          95.96%
                          11

Length: 13079
Matches: 216
Conservative: 3
Mismatches: 6
Indels: 0
Gaps: 0

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly   20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  186 ATGGCGTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  245

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
        ::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  246 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  305

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
        |||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db  306 ACCCAGAACACGGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  365

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  366 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  425

Qy   81 AspThrTyrLysLeuSerPhePheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
        |||||||||||||||||||||||||||||||||||||||||||||
Db  426 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  485

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
        ||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  486 GACGGCGGCGTGGTGACCGTGACCCAGGACTCTCCCTGCAGGACGGCTGCTTCATCTAC  545

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  546 AAGGTGAAGTTCATCGGGGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT  605

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  606 ATGGGCTGGGAGGCCAGCCCCTCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  665

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  666 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  725

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  726 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  785
```

RESULT 63
AF506025

```
LOCUS       AF506025                 681 bp    DNA     linear   SYN 19-JUN-2002
DEFINITION  Synthetic construct dimeric red fluorescent protein gene, complete
            cds.
ACCESSION   AF506025
VERSION     AF506025.1  GI:21464833
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S.,
            Zacharias,D.A. and Tsien,R.Y.
  TITLE     A monomeric red fluorescent protein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7877-7882 (2002)
  PUBMED    12060735
REFERENCE   2  (bases 1 to 681)
  AUTHORS   Campbell,R.E. and Tsien,R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La
            Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1..681
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="from Discosoma sp.; Clontech vector pDsRed-N1 with
                     human optimized codon usage"
                     1..681
                     /note="dimer2; engineered variant of tetrameric red
                     fluorescent protein from Discosoma deposited in GenBank
                     Accession Number AAF03369"
     CDS             1..681
                     /codon_start=1
                     /transl_table=11
                     /product="dimeric red fluorescent protein"
                     /protein_id="AAM54542.1"
                     /db_xref="GI:21464834"
                     /translation="MVASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQT
     AKLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFED
     GGVVTVTQDSSLQDGTLIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKG
     EIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERS
     EGRHHLFL"
```

ORIGIN

Alignment Scores:
Pred. No.:
Score: 4.42e-105
Percent Similarity: 1121.00
Best Local Similarity: 95.96%
Query Match: 92.83%
                          92.34%
                          11

Length: 681
Matches: 207
Conservative: 7
Mismatches: 9
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AF506025 (1-681)

```
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlnTyrGluArgThrGluGlyArg  220
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  786 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGC  845

Qy  221 HisHisLeuPheLeu  225
        |||||||||||||
Db  846 CACCACCTGTTCCTG  860

Qy    3 SerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrVal   22
        |||||||||||||||:|||||||||||||||||||||||||||:|||
Db   10 TCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTG   69

Qy   23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn   42
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCAG  129

Qy   43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer   62
```

```
RESULT 64

LOCUS       AF506026                1395 bp    DNA     linear   SYN 19-JUN-2002
DEFINITION  Synthetic construct tandem-dimer red fluorescent protein gene,
            complete cds.
ACCESSION   AF506026
VERSION     AF506026.1  GI:21464835
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1395)
  AUTHORS   Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S.,
            Zacharias,D.A. and Tsien,R.Y.
  TITLE     A monomeric red fluorescent protein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7877-7882 (2002)
   PUBMED   12060735
REFERENCE   2  (bases 1 to 1395)
  AUTHORS   Campbell,R.E. and Tsien,R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La
            Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1..1395
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="from Discosoma sp.; Clontech vector pDsRed-N1 with
                     human optimized codon usage"
     CDS             1..1395
                     /note="cdimer2(12); engineered variant of tetrameric red
                     fluorescent protein from Discosoma deposited in GenBank
                     Accession Number AAF03369"
                     /codon_start=1
                     /transl_table=11
                     /product="tandem-dimer red fluorescent protein"
                     /protein_id="AAM54543.1"
                     /db_xref="GI:21464836"
                     /translation="MVASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGRPYEGTQT
                     AKLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFED
                     GGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERLYPRDGVLKG
                     EIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERS
                     EGRHHLFLGHGTGSTGSGSSGTASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGRP
                     YEGTQTAKLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYKKLSFPEGFKWER
                     VMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERLYPR
                     DGVLKGEIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNEDYTIV
                     EQYERSEGRHHLFL"
ORIGIN
```

US-10-006-922A-12 (1-225) x AF506026 (1-1395)

Alignment Scores:
Pred. No.:
Score: 1121.00
Percent Similarity: 95.96%
Best Local Similarity: 92.83%
Query Match: 92.34%
DB: 11

Length: 1395
Matches: 207
Conservative: 7
Mismatches: 9
Indels: 0
Gaps: 0

```
Qy          3 SerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGlyThrVal 22
              ::|||:|:||||||||||||||||:|||||:::||||||:::|||
Db         10 TCCTCCGAGGACGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTG 69

Qy         23 AsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGlyThrGlnHisAsn 42
              |||||||:|||||||||||||||||||||:|||||||||||||:::||||||
Db         70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGGCGCCCCTACGAGGGCACCCAG 129

Qy         43 ThrValLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
              |::|||||||||:::||||||||||||||::||||:::|||||
Db        130 ACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 189

Qy         63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        190 CCCCAGTTCCAGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGACTAC 249

Qy         83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 309

Qy        103 GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal 122
              ||||||||||||||||||||||||||||||||||||||||||||||::||||||||||
Db        310 GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG 369

Qy        123 LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 AAGTTCCGCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC 429

Qy        143 TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis 162
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC 489

Qy        163 LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet 182
              |||||||||||||||||||||||||||||||||||:::||||||||||
Db        490 CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATG 549

Qy        183 AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr 202
              ||||||||||||||||||||:::||||||||||||||||||||||||:::||||||
Db        550 GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACACCAAGCTGGACATCACC 609

Qy        203 SerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHis 222
              ||||||||||||||||||||||||||||||||::|||||||||||||||
Db        610 TCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGGTCCGAGGGCCGCCACCAC 669

Qy        223 LeuPheLeu 225
              |||||||||
Db        670 CTGTTCCTG 678
```

RESULT 65

```
LOCUS           AX824729                  678 bp    DNA     linear   PAT 11-DEC-2003
DEFINITION      Sequence 11 from Patent WO02068459.
ACCESSION       AX824729
VERSION         AX824729.1  GI:39750593
KEYWORDS        .
SOURCE          synthetic construct
  ORGANISM      synthetic construct
                other sequences; artificial sequences.
REFERENCE    1  (bases 1 to 678)
  AUTHORS
  TITLE         Non aggregating fluorescent proteins and methods for using the same
  JOURNAL       Patent: WO 02068459-A1 11 06-SEP-2002;
FEATURES             Location/Qualifiers
     source      1..678
                 /organism="synthetic construct"
                 /mol_type="unassigned DNA"
                 /db_xref="taxon:32630"
                 /note="hybrid coding sequence"
```

ORIGIN

US-10-006-922A-12 (1-225) x AX824729 (1-678)

Alignment Scores:

```
Pred. No.:
Score:               1119.00         Length:        678
Percent Similarity:  96.89%          Matches:       206
Percent Similarity:  91.56%          Conservative:  12
Best Local Similarity: 92.17%        Mismatches:    7
Query Match:         6               Indels:        0
DB:                                  Gaps:          0
```

```
Qy    1 MetArgSerSerLysAsnValIleLysGluGluGlnPheMetArgPheLysValAlaArgMetGluGly 20
          |||
Db    1 ATGAGCTGCAGCAAGAACGTGATCAAGGAGGAGTTCATGCGGTTCAAGGTGCCAGTGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||::||||||||||||::||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGGCACGAGTTCGAGATCAAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||::|||||||||||||||||||||||||||||||||
Db  121 CACTGCAGCGTGAAGCTCATGGTGACCAAGGGCGGGCCCCTCCCCTTCGCCTTCGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysIleTyrValLysHisProAlaAspIlePro 80
          ::|||||||||||||||||||||||||||||||||||||||||||||||||||::|
Db  181 CTCAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCGACATCCC 240

Qy   81 AspThrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||::|||||||||||||||||||||||||||||||||||
Db  241 GACTACAAGAAGCTTCAGCTTCCCGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||
Db  301 GACGGCGGCGTGACCGTGACCCAGGACAGCAGCCTGCTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          :::||||||||||||||||||||||||||||::|||||||||::||||
Db  361 GAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||::|||||||||||||||
Db  421 CGGGGCCGTGGGAGGCCAGCACGGAGCGGCTCTACCCCCGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisHisLysLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ::|||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATCCACAACAAGCTCCGGCTCGAGGGCGGGGGCCACTACCTCGTGGAGTTCAAGAGCATC 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ::|||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACAGCAAGCTCGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 ... 600
```

RESULT 66

```
LOCUS           AY678268                  705 bp    DNA     linear   SYN 17-DEC-2004
DEFINITION      Synthetic construct dimeric red fluorescent protein gene, complete
                cds.
ACCESSION       AY678268
VERSION         AY678268.1  GI:55420618
KEYWORDS        .
SOURCE          synthetic construct
  ORGANISM      synthetic construct
                other sequences; artificial sequences.
REFERENCE    1  (bases 1 to 705)
  AUTHORS       Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
                Palmer,A.E. and Tsien,R.Y.
  TITLE         Improved monomeric red, orange and yellow fluorescent proteins
                derived from Discosoma sp. red fluorescent protein
  JOURNAL       Nat. Biotechnol. 22 (12), 1567-1572 (2004)
  PUBMED        15558047
REFERENCE    2  (bases 1 to 705)
  AUTHORS       Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
                Tsien,R.Y.
  TITLE         Direct Submission
  JOURNAL       Submitted (06-JUL-2004) Pharmacology, University of California, San
                Diego, 9500 Gilman Dr.; La Jolla, CA 92093-0647, USA
FEATURES             Location/Qualifiers
     source      1..705
                 /organism="synthetic construct"
                 /mol_type="other DNA"
                 /db_xref="taxon:32630"
     CDS         1..705
                 /note="dTomato; engineered variant of dimeric red
                 fluorescent protein dimer2 in GenBank Accession Number
                 AF506025; N- and C-termini replaced with equivalent
                 residues from EGFP"
                 /codon_start=1
                 /transl_table=11
                 /product="dimeric red fluorescent protein"
                 /db_xref="GI:55420619"
                 /protein_id="AAV52168.1"
                 /translation="MVSKGEEVIKEMRFKVRMEGSMNGHEFEIEGEGEGRPYEGTQT
                 AKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFED
                 GGLVTVTQDSSLQDGTLIYKVKMRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKG
                 EIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERS
                 EGRHHLFLYGMDELYK"
```

ORIGIN

US-10-006-922A-12 (1-225) x AY678268 (1-705)

Alignment Scores:

```
Pred. No.:
Score:               1095.00         Length:        705
Percent Similarity:  94.57%          Matches:       201
Percent Similarity:  90.95%          Conservative:  8
Best Local Similarity: 90.20%        Mismatches:    12
Query Match:         11              Indels:        0
DB:                                  Gaps:          0
```

```
Qy    5 LysAsnValIleLysGluGluGlnPheMetArgPheLysValAlaArgMetGluGlyThrValAsnGly 24
          |:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5 GAGGAAGTTCATCAAAGAGTTCATGCGCTTCAAGGTGCGTATGGAGGGCTCCATGAACGGC 75

Qy   25 HisGluPheIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal 44
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   76 CACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCCAGACCGCC 135

Qy   45 LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln 64
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  ... 
```

```
Db  601 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGAGCGAGGGCCGG 660
Qy  221 HisHisLeuPheLeu 225
          |||||||||||
Db  661 CACCACCTTCTTCCTC 675
```

RESULT 67

LOCUS   AY678269   1431 bp   DNA   linear   SYN 17-DEC-2004
DEFINITION   Synthetic construct tandem-dimer red fluorescent protein gene,
                complete cds.
ACCESSION   AY678269
VERSION   AY678269.1 GI:55420622
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM   synthetic construct
                other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 1431)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
                Palmer,A.E. and Tsien,R.Y.
  TITLE   Improved monomeric red, orange and yellow fluorescent proteins
                derived from Discosoma sp. red fluorescent protein
  JOURNAL   Nat. Biotechnol. 22 (12), 1567-1572 (2004)
   PUBMED   15558047
REFERENCE   2 (bases 1 to 1431)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
                Tsien,R.Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Pharmacology, University of California, San
                Diego, 9500 Gilman Dr; La Jolla, CA 92093-0647, USA
FEATURES   Location/Qualifiers
     source   1..1431
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS   1..1431
                     /note="cdTomato; engineered variant of dimeric red
                     fluorescent protein dimer2 in GenBank Accession Number
                     AF506025; N- and C-termini replaced with equivalent
                     residues from EGFP"
                     /codon_start=1
                     /transl_table=11
                     /product="tandem-dimer red fluorescent protein"
                     /protein_id="AAV52169.1"
                     /db_xref="GI:55420623"
                     /translation="MVSKGEEVIKEFMRFKVRMEGSMNGHEFEIEGEGRPYEGTQT
                     AKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFED
                     GGLVTVTQDSSLQDGTLIYKVKMRGTNFPPDGPVMQKKTMGWEASTERLYPRDGVLKG
                     EIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERS
                     EGRHHLFLGHGTGSTGSGSSGTASSEDNNMAVIKEFMRFKVRMEGSMNGHEFEIEGEG
                     EGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYKKLSFPEGF
                     KWERVMNFEDGGLVTVTQDSSLQDGTLIYKVKMRGTNFPPDGPVMQKKTMGWEASTER
                     LYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNED
                     YTIVEQYERSEGRHHLFLYGMDELYK"
ORIGIN

Alignment Scores:
Pred. No.:   11
Score: 1095.00
Percent Similarity: 94.57%
Best Local Similarity: 90.20%
Query Match: 11

Length: 1431
Matches: 201
Conservative: 8
Mismatches: 12
Indels: 0
Gaps: 0

DB:

US-10-006-922A-12 (1-225) x AY678269 (1-1431)

```
Qy      5 LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly  24
            :::|  |||  ||  || || || |||  || || |||||  |||    |  |
Db     16 GAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGC  75

Qy     25 HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal  44
           |||||||||||||||||||||||||||||||||||||||||||||    ||  ||
Db     76 CACGAGTTCGAGATCGAAGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCCAGACGCC  135

Qy     45 LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln  64
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAG  195

Qy     65 PheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLys  84
           |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    196 TTCATGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGATTACAAGAAG  255

Qy     85 LeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyVal  104
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db    256 CTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTC  315

Qy    105 ValThrValThrGlnAspSerSerLeuGlnAspGlyThrLeuIleTyrLysValLysPhe  124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db    316 GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGATG  375

Qy    125 IleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu  144
           |||  | ||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db    376 CGCGGCGCCAACTTCCCCCCCGACGGCCCCGTAATCCAGAAGAAGACCATGGGCTGGGAG  435

Qy    145 AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuGlyGluIleHisLysAla  164
           |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db    436 GCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTCCTGAAGGGCGAGATCCACAAGGCC  495

Qy    165 LeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys  184
           |||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db    496 CTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAG  555

Qy    185 LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis  204
           |||||||||||||||||||||||||||||||||||||:::|||  || || ||||||||
Db    556 AAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACCTCCCAC  615

Qy    205 AsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHisLeuPhe  224
           |||||||||||||||||||||||||||||:::|||  || |||||||||||||||||||
Db    616 AACGAGGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTC  675

Qy    225 Leu 225
           |||
Db    676 CTG 678
```

RESULT 68
AF272711

```
LOCUS       AF272711      876 bp    mRNA    linear   INV 26-SEP-2000
DEFINITION  Discosoma sp. SSAL-2000 red fluorescent protein (FPS93) mRNA,
            complete cds.
ACCESSION   AF272711
VERSION     AF272711.1  GI:10304306
KEYWORDS    .
SOURCE      Discosoma sp. SSAL-2000
  ORGANISM  Discosoma sp. SSAL-2000
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   Fradkov,A.F., Chen,Y., Ding,L., Barsova,E.V., Matz,M.V. and
            Lukyanov,S.A.
  TITLE     Novel fluorescent protein from Discosoma coral and its mutants
            possesses a unique far-red fluorescence
  JOURNAL   FEBS Lett. 479 (3), 127-130 (2000)
   PUBMED   10981720
REFERENCE   2  (bases 1 to 876)
  AUTHORS   Fradkov,A.F., Chen,Y., Ding,L., Barsova,E.V., Matz,M.V. and
            Lukyanov,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2000) Institute of Bioorganic Chemistry RAS,
            Miklukho-Maklaya 16/10, Moscow 117871, Russia
FEATURES             Location/Qualifiers
     source          1..876
                     /organism="Discosoma sp. SSAL-2000"
                     /mol_type="mRNA"
                     /db_xref="taxon:137428"
     gene            1..876
                     /gene="FPS93"
     CDS             45..737
                     /gene="FPS93"
                     /note="dsFP593"
                     /codon_start=1
                     /product="red fluorescent protein"
                     /protein_id="AAG16224.1"
                     /db_xref="GI:10304307"
                     /translation="MSCSKNVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSV
                     KLMVTKGGPLPFAPDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDG
                     GVVTVSQDSSLKDGCFIYEVKFIGVNFPSDGPVMQRRTKGWEASSERLYPRDGVLKGD
                     IHMALRLEGGGHYLVEFKSIYMVKKPSVQLPGYYYVVDSKLDMTSHNEDYTVVEQYEKT
                     QGRHHPIKPLQ"
ORIGIN
```

US-10-006-922A-12 (1-225) x AF272711 (1-876)

```
Alignment Scores:
Pred. No.:                2.53e-101      Length:        876
Score:                    1085.50        Matches:       199
Percent Similarity:       95.58%         Conservative:   17
Best Local Similarity:    88.05%         Mismatches:      9
Query Match:              89.42%         Indels:          1
DB:                       2              Gaps:            1
```

```
Qy    1 MetArgSerIleLysSerAlaAsnValIleLeuSGlyGlyProLeuProPheAlaTrpAspIle 20
          |||  ||||||||||||||||||||||||||||||||||||||||||||
Db    1 MetArgSerSerIleLysSerAlaAsnValIleLeuSGlyGlyProLeuProPheAlaTrpAspIle 20

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   45 ATGAGTTGTTCCAAGAATGTGATCAAGGAGTTCATGAGGTTCGTATGGGAAGGA 104

Qy   41 HisAsnThrValLysSerLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          || | :::   ||||||||||| ::| ||||||||||||||||||||
Db  165 CACTGTTCCGTAAAGCTTATGGTAACCAAGGGGTGGACCTTGCCATTGCTTTTGATATT 224

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  225 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAACACCCTGCCGACATACCA 284

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  285 GACTATAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 344
```

RESULT 69
AX686894

```
LOCUS       AX686894      876 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 17 from Patent WO0127150.
ACCESSION   AX686894
VERSION     AX686894.1  GI:29409472
KEYWORDS    .
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Lukyanov,S.A., Fradkov,A.F., Labas,Y.A., Matz,M.V. and Terskikh,A.
  TITLE     Anthozoa derived chromo/fluoroproteins and methods for using the
            same
  JOURNAL   Patent: WO 0127150-A 17 19-APR-2001;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1..876
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"
ORIGIN
```

US-10-006-922A-12 (1-225) x AX686894 (1-876)

```
Alignment Scores:
Pred. No.:                2.53e-101      Length:        876
Score:                    1085.50        Matches:       199
Percent Similarity:       95.58%         Conservative:   17
Best Local Similarity:    88.05%         Mismatches:      9
Query Match:              89.42%         Indels:          1
DB:                       6              Gaps:            1
```

```
Qy    1 MetArgSerSerIleLysSerAlaAsnValIleLeuGluLysMetGlyGluPheMetArgPheLysValAlaArgMetGluGly 20
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    1 MetArgSerSerIleLysSerAlaAsnValIleLeuGluLysMetGlyGluPheMetArgPheLysValAlaArgMetGluGly 20

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   45 ATGAGTTGTTCCAAGAATGTGATCAAGGAGTTCATGAGGTTCGTATGGGAAGGA 104

Qy  105 ACGGTCAATGGGCACGAGTTTGAAATAAAAGGCGAAGGTGAAGGGAGGCCTTACGGAGGT 164
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105 ACGGTCAATGGGCACGAGTTTGAAATAAAAGGCGAAGGTGAAGGGAGGCCTTACGGAGGT 164
```

RESULT 70

```
Qy   41 HisAsnThrValLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||  ::|||  |||||  |||||||||||||||||||||||||
Db  165 CACTCGTTCCGTAAAGCTTATGGTAGTAAACAAGGGTGGACCTTTGCCATTTGGATAATT 224

Qy   61 LeuSerProGlnPheGlnTyrGlySerLeuSerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||  ||||||     :::  |||  |||||||  |||
Db  225 TTGTCACCAAATTTCAGTATGGGAGCAAGGTATGTATGTCAAACACCCTGCGACATACCA 284

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  285 GACTATAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 344

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||  |||||||||||||||
Db  345 GACGGTGGCGTGGTTACTGTAACCCAAGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 404

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        :::  |||||||||||  |||||||||||||||:::::::|||  |||
Db  405 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCCGTTATGCAGAAGAAGACA 464

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||  |||||||||||||  :::  |||||||||||  |||||||||:::
Db  465 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGCTGAAGGGAGAC 524

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||  |||||||  |||||||||||||||:::  |||||:::  |||
Db  525 ATCCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTCGTTGAATTCAAAAGTATT 584

Qy  181 TyrMetAlaLysLysPro---ValGlnLeuProGlyTyrTyrTyrTyrTyrValAspSerLysLeu 199
        |||:::||||||||||     |||||||||||||||||||||||||:::
Db  585 TACATGGTAAAGAAGCCTTCAGTGCAGTTGCCAGGCTACTATTATTATTATGTTGACTCCAAACTG 644

Qy  200 AspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGly 219
        |||  |||||||||||||  |||||||||||||||||||||||  |||
Db  645 GATATGACGAGCCACAACGAAGATTACACAGTCGTTGAGCAGTATGAAAAAACCCAGGGA 704

Qy  220 ArgHisHisLeuPheLeu 225
        |||||||||  |::::
Db  705 CGCCACCACCATCCGTTCATT 722
```

LOCUS   AF506027   678 bp   DNA   linear   SYN 19-JUN-2002

DEFINITION Synthetic construct monomeric red fluorescent protein gene, complete cds.

ACCESSION   AF506027

VERSION   AF506027.1 GI:21464837

KEYWORDS   .

SOURCE   synthetic construct

  ORGANISM synthetic construct
          other sequences; artificial sequences.

REFERENCE   1 (bases 1 to 678)

  AUTHORS   Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S., Zacharias,D.A. and Tsien,R.Y.

  TITLE   A monomeric red fluorescent protein

  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7877-7882 (2002)

  PUBMED   12060735

REFERENCE   2 (bases 1 to 678)

  AUTHORS   Campbell,R.E. and Tsien,R.Y.

  TITLE   Direct Submission

  JOURNAL   Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA

FEATURES   Location/Qualifiers

     source   1..678
                 /organism="synthetic construct"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:32630"
                 /note="synthetic construct"

     CDS   1..678
                 /note="from Discosoma sp.; Clontech vector pDsRed-N1 with human optimized codon usage"
                 /note="mRFP1; engineered variant of tetrameric red fluorescent protein from Discosoma deposited in GenBank Accession Number AAF03369"
                 /codon_start=1
                 /transl_table=11
                 /product="monomeric red fluorescent protein"
                 /protein_id="AAM54544.1"
                 /db_xref="GI:21464838"
                 /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAE
GRHSTGA"

ORIGIN

US-10-006-922A-12 (1-225) x AF506027 (1-678)

Alignment Scores:

Pred. No.:	Length:	678
Score: 1021.00	Matches:	192
Percent Similarity: 89.14%	Conservative:	5
Best Local Similarity: 86.88%	Mismatches:	24
Query Match: 84.10%	Indels:	0
DB:	Gaps:	0

7.14e-95
11

```
Qy    1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        |||:::  :::|||  :::||||||||||||||||||||||||||
Db    1 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        :::|||||||||||||||||||||||||||||||||||||||||||
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||  |||||||||||||||||||||||||||||||  |||||||
Db  121 CACCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysAlaTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||:::|||||||||||||||||||
Db  181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||:::||||||||||
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        :::|||||||  ||||||||||||||||||||||||  :::|||||||||
Db  361 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||  |||||||||||||  :::||||||  |||||||||  |||||||||:::
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGGTATCCCCGAGGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||:::|||||||  |||||||||||||||||:::||||||:::  |||
Db  481 ATCAAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGAGGCCGAAGTCAAGACCACC 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrTyrValAspSerLeuAsp 200
        |||:::||||||||||||||||||||||||||:::  ||||||:::
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||  |||||||||||||  |||||||||||||||||||  :::|||||||
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC 660

Qy  221 His 221
        |||
Db  661 CAC 663
```

RESULT 71

LOCUS   AB166761   696 bp   DNA   linear   SYN 02-OCT-2004

```
DEFINITION  Synthetic construct monomeric DsRed variant gene for red
            fluorescent protein, partial cds.
ACCESSION   AB166761
VERSION     AB166761.1  GI:53748885
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Matsuoka,K.
  TITLE     Monomeric DsRed variant with plant codon usage
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 696)
  AUTHORS   Matsuoka,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2004) Ken Matsuoka, RIKEN (The Institute of
            Physical and Chemical Research), Plant Science Center; 1-3-22
            Suehirocho, Tsurumi-ku, Yokohama 230-0045, Japan
            (E-mail:kenmat@psc.riken.jp,
            URL:http://mrg.psc.riken.go.jp/strc/STRC%20Lab%20index.htm,
            Tel:81-45-503-9575, Fax:81-45-503-9573)
FEATURES             Location/Qualifiers
     source          1..696
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     gene            1..696
                     /gene="monomeric DsRed variant"
     CDS             8..>696
                     /gene="monomeric DsRed variant"
                     /codon_start=1
                     /transl_table=11
                     /product="red fluorescent protein"
                     /protein_id="BAD52341.1"
                     /db_xref="GI:53748886"
                     /translation="MASSEDVIKEPMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAE
                     GRHSTGARPPES"
ORIGIN
```

Alignment Scores:
```
Pred. No.:
Score:                7.37e-95
Score:                1021.00
Percent Similarity:   89.14%
Percent Identity:     86.88%
Best Local Similarity: 84.10%
Query Match:          11
```

US-10-006-922A-12 (1-225) x AB166761 (1-696)

```
Length:        696
Matches:       192
Conservative:    5
Mismatches:     24
Indels:          0
Gaps:            0
```

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20

Qy    1 ATGGCATCTTCCGAGGATGTGATTAAGGAGTTTATGAGATTCAAGGTGAGGATGGAGGGA 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8 ATGGCATCTTCCGAGGATGTGATTAAGGAGTTTATGAGATTCAAGGTGAGGATGGAGGGA 67

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 TCTGTTAATGGACATGAGTTCGAGATTGAGGGCGAGGAGGGAAGAGACCTTACGAGGGG 127

Qy   41 HisGluAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  128 ACTCAGACTGCTAAGCTGAAGGTTACTAAGGGAGGACCACTGCCTTTCGCTTGGGATATC 187

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||
Db  188 CTGTCTCCACAATTCCAATACGGATCTAAGGTCTACGTCAAGCATCCAGCAGATATCCCA 247

Qy   81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||
Db  248 GACTATTTGAAGCTCTCTTTCCCAGAAGGATTCAAGTGGGAGAGAGTTATGAACTTCGAG 307
```

RESULT 72
```
LOCUS       DQ005472       4543 bp    DNA     circular SYN 07-MAY-2005
DEFINITION  Cloning vector pSAT6A-RFP-N1, complete sequence.
ACCESSION   DQ005472
VERSION     DQ005472.1  GI:63002541
KEYWORDS    .
SOURCE      Cloning vector pSAT6A-RFP-N1
  ORGANISM  Cloning vector pSAT6A-RFP-N1
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4543)
  AUTHORS   Chung,S.-M., Vyas,S. and Tzfira,T.
  TITLE     A versatile vector system for multiple gene expression in plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4543)
  AUTHORS   Chung,S.-M., Vyas,S., Kozlovsky,S., Citovsky,V. and Tzfira,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2005) Department of Biochemistry and Cell
            Biology, State University of New York at Stony Brook, Stony Brook,
            NY 11794, USA
FEATURES             Location/Qualifiers
     source          1..4543
                     /organism="Cloning vector pSAT6A-RFP-N1"
                     /mol_type="other DNA"
                     /db_xref="taxon:326515"
     promoter        441..762
                     /note="CaMV 35S promoter"
     promoter        768..1089
                     /note="CaMV 35S promoter"
     enhancer        1190..1320
                     /note="translational enhancer 5'-UTR from tobacco etch
                     virus"
     misc_feature    1321..1379
                     /note="multiple cloning site; MCS"
     CDS             1380..2057
                     /codon_start=1
                     /product="RFP"
                     /protein_id="AAY25379.1"
                     /db_xref="GI:63002542"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAE
                     GRHSTGA"
```

```
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308 GATGGAGGAGTTGTTACTGTTACTCAAGATTCCTCCCTTCAAGATGGGTGCTTCATCTAC 367

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  368 AAGGTGAAGCTCAGGAGAACTAACTTCCCATCTGATGGACCAGTTATGCAAAAGAAGACT 427

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||:::||||||||||||||||||||||||||||||
Db  428 ATGGGATGGGAGGCGAGCACCGAACGTCTTTCTTCCACGATGGTGCTCTCAAGGGAGAG 487

Qy  161 IleHisGlyAlaLeuGlnLeuGluGlyHisGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  488 ATCAAGATGAGAGACTCAAGCTGGAAGAGTGGACANCATTNTGAGGTCGAAGTCAAGA 547

Qy  181 TyrMetAlaLysLysProGlnGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||
Db  548 TACATGGCTAAGAAGCCCAGTCAAACTTCCAGGATACTACGAGACTCAAGAGCTCGAT 607

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||:::
Db  608 ATCACCTCCCATAACGAGGATTACACCATCGTTGGAACAAATATGAGGAGGGCTGAAGG 667

Qy  221 His 221
        |||
Db  668 CAT 670
```

Left column

```
ORIGIN
terminator   2064..2274
             /note="CaMV 35S terminator"
CDS          complement(3483..4343)
             /note="amp resistance"
             /codon_start=1
             /product="Bla"
             /protein_id="AAY25380.1"
             /db_xref="GI:63002543"
             /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
             IELDLNGCKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSONDLVE
             YSPVTEKHLTDGMTVRELCSAATMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
             DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
             LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVIYTTGSQATMDERNRQIA
             EIGASLIKHW"
```

US-10-006-922A-12 (1-225) x DQ005472 (1-4543)

Alignment Scores:
Pred. No.:

Score:	7.26e-94	Length: 4543
Percent Similarity:	1021.00	Matches: 192
Best Local Similarity:	89.14%	Conservative: 5
Query Match:	86.88%	Mismatches: 24
DB:	84.10%	Indels: 0
	11	Gaps: 0

```
Qy     1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1380 ATGGCCTCCTTCGAGGACGTTCATCAAGGAGTTCATGCGCTTCAAGGTGCCGATGGAGGGC 1439

Qy    21 ThrValAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1440 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1499

Qy    41 HisAsnThrValLeuValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1500 ACCCAGACCGGCCAAGCTGAAGGTGAACAAGGGCGCCCCCCTGCCCTTCGCCTGGGACATC 1559

Qy    61 LeuSerProGlnGlnMetTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1560 CTGTCCCCCCAGTTCCAGTTCACGGGCTTCCAAGGCGTGAGCGACATCCCC 1619

Qy    81 AspGlyTyrLysLysLeuSerPheProGluGlyPheThrTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1620 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1679

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1680 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1739

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1740 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 1799

Qy   141 MetGlyTrpGluAlaSerThrGluArgMetTyrProGluAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1800 ATGGGCTGGGAGGCCTCCACCGAGCGCATGTACCCCGAGGACGGCGTCCTGAAGGGCGAG 1859

Qy   161 IleLysHisLysLeuLysGlyTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||:::
Db  1860 ATCAAGCATAAGCTGAAGGGCTACCTGGTCGAGTTCAAGAGCATC 1919

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1920 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 1979

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1980 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGC 2039

Qy   221 His 221
         |||
Db  2040 CAC 2042
```

Right column

RESULT 73

```
DQ005474
LOCUS       DQ005474    4555 bp    DNA    circular  SYN  07-MAY-2005
DEFINITION  Cloning vector pSAT6-RFP-C1, complete sequence.
ACCESSION   DQ005474
VERSION     DQ005474.1  GI:63002546
KEYWORDS    .
SOURCE      Cloning vector pSAT6-RFP-C1
  ORGANISM  Cloning vector pSAT6-RFP-C1
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4555)
  AUTHORS   Chung,S.-M., Vyas,S., Kozlovskiy,S., Citovsky,V. and Tzfira,T.
  TITLE     A versatile vector system for multiple gene expression in plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4555)
  AUTHORS   Chung,S.-M., Vyas,S. and Tzfira,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2005) Department of Biochemistry and Cell
            Biology, State University of New York at Stony Brook, Stony Brook,
            NY 11794, USA
FEATURES             Location/Qualifiers
     source          1..4555
                     /organism="Cloning vector pSAT6-RFP-C1"
                     /mol_type="other DNA"
                     /db_xref="taxon:326517"
     promoter        441..762
                     /note="CaMV 35S promoter"
     promoter        768..1089
                     /note="CaMV 35S promoter"
     enhancer        1190..1320
                     /note="CaMV 35S promoter"
                     /note="translational enhancer 5'-UTR from tobacco etch
                     virus"
     CDS             1323..>1997
                     /codon_start=1
                     /product="RFP"
                     /protein_id="AAY25382.1"
                     /db_xref="GI:63002547"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAE
                     GRHSTGA"
     misc_feature    1998..2075
                     /note="multiple cloning site; MCS"
     terminator      2076..2286
                     /note="CaMV 35S terminator"
     CDS             complement(3495..4355)
                     /note="amp resistance"
                     /codon_start=1
                     /product="Bla"
                     /protein_id="AAY25383.1"
                     /db_xref="GI:63002548"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAATMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
ORIGIN
```

Alignment Scores:
Pred. No.:

Score:	7.28e-94	Length: 4555
Percent Similarity:	1021.00	Matches: 192
Best Local Similarity:	89.14%	Conservative: 5
Query Match:	86.88%	Mismatches: 24
DB:	84.10%	Indels: 0
	11	Gaps: 0

US-10-006-922A-12 (1-225) x DQ005474 (1-4555)

```
Qy  1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
      ||||||||||||||||:::::::||||||||||||||||||||||||||||||||||||
```

```
RESULT 74

LOCUS       DQ005475               4558 bp    DNA     circular SYN 07-MAY-2005
DEFINITION  Cloning vector pSAT6-RFP-N1, complete sequence.
ACCESSION   DQ005475
VERSION     DQ005475.1  GI:63002549
KEYWORDS    .
SOURCE      Cloning vector pSAT6-RFP-N1
  ORGANISM  Cloning vector pSAT6-RFP-N1
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4558)
  AUTHORS   Chung,S.-M., Vyas,S. and Tzfira,T.
  TITLE     A versatile vector system for multiple gene expression in plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4558)
  AUTHORS   Chung,S.-M., Vyas,S., Kozlovskiy,S., Citovsky,V. and Tzfira,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2005) Department of Biochemistry and Cell
            Biology, State University of New York at Stony Brook, Stony Brook,
            NY 11794, USA
FEATURES             Location/Qualifiers
     source          1..4558
                     /organism="Cloning vector pSAT6-RFP-N1"
                     /mol_type="other DNA"
                     /db_xref="taxon:326518"
     promoter        441..762
                     /note="CaMV 35S promoter"
     promoter        768..1089
                     /note="CaMV 35S promoter"
     enhancer        1190..1320
                     /note="translational enhancer 5'-UTR from tobacco etch
                     virus"
     misc_feature    1321..1394
                     /note="multiple cloning site; MCS"
     CDS             1395..2072
                     /codon_start=1
                     /product="RFP"
                     /protein_id="AAY25384.1"
                     /db_xref="GI:63002550"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
                     GVVTVQDDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAE
                     GRHSTGA"
     terminator      2079..2289
                     /note="CaMV 35S terminator"
     CDS             complement(3498..4358)
                     /note="amp resistance"
                     /codon_start=1
                     /product="Bla"
                     /protein_id="AAY25385.1"
                     /db_xref="GI:63002551"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
ORIGIN
```

US-10-006-922A-12 (1-225) x DQ005475 (1-4558)

Alignment Scores:

Pred. No.:	1	
Score:	1021.00	Length:	4558
Percent Similarity:	89.14%	Matches:	192
Percent Identity:	86.88%	Conservative:	5
Best Local Similarity:	84.41%	Mismatches:	24
Query Match:	11	Indels:	0
DB:		Gaps:	0

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||::::||::||:::||
Db  1395 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1454

Qy    21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1455 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1514

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1515 ACCCAGACGGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1574

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysAlaTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1575 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 1634

Qy    81 AspTyrLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1635 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1694

Qy   101 AspGlyGlyValValThrValGlnAspAspSerSerLeuGlnAspGlyGluPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1695 GACGGCGGCGTGGTGACCGTGCAGGACGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 1754

Qy   121 LysValLysLeuArgGlyThrAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1755 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 1814

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||
Db  1815 ...
```

RESULT 75

```
LOCUS       AJ851287     5311 bp    DNA     circular SYN 12-NOV-2004
DEFINITION  Cloning vector pRU1144 mRFP1 gene for mono Red fluorescent protein.
ACCESSION   AJ851287
VERSION     AJ851287.1  GI:55724880
KEYWORDS    mono Red fluorescent protein; mRFP1 gene.
SOURCE      Cloning vector pRU1144
  ORGANISM  Cloning vector pRU1144
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE     High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5311)
  AUTHORS   Poole,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..5311
                     /organism="Cloning vector pRU1144"
                     /mol_type="other DNA"
                     /db_xref="taxon:299182"
                     /lab_host="Escherichia coli"
     gene            326..1003
                     /gene="mRFP1"
     CDS             326..1003
                     /gene="mRFP1"
                     /codon_start=1
                     /transl_table=11
                     /product="mono Red fluorescent protein"
                     /protein_id="CAH64892.1"
                     /db_xref="GI:55724881"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAE
                     GRHSTGA"
```

```
Qy       1 MetArgSerArgLysArgAsnValIleGluPheMetArgPheLysValArgMetGluGly 20
           |||||:||||||||||:|:||||||||||||||||||||||||||||||||||||||||
Db     326 ATGGCGCTCCGGAGGACGTCATCGAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 385
```

```
Qy      21 ThrValAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           :::|||||||||||||||||:|||||:||||||||||||||||||||||||||||||||||
Db     386 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 445
```

```
Qy      41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     446 ACCCAGACACCGCACCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 505
```

```
Qy      61 LeuSerProGlnPheGlnTyrGlySerLysAlaTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     506 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 565
```

```
Qy      81 AspTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     566 GACTACCTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 625
```

```
Qy     101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     626 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 685
```

```
Qy     161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrAspAlaGluValLysThrTyrLeuVal 180
           |||||||||||||||||:||::||||:::
Db    1815 ATGGGGCTGGGAGGCGGCTCCACCGGAGCTTGA 1874
```

```
Qy     161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrAspAlaGluValLysThrTyrLeuVal 180
           |||||||||||||||||:|:||
Db    1875 ATCAAGATGAGGCTGAAGCTGAAAGACGGGCCACTACGACGCCGAGGTCAAGACCTAC 1934
```

```
Qy     181 TyrMetAlaLysLysProValGlnLeuProGlyAlaTyrLysThrAspIleLysLeuAsp 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1935 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGAC 1994
```

```
Qy     201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1995 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC 2054
```

```
Qy     221 His 221
           |||
Db    2055 CAC 2057
```

RESULT 76

```
LOCUS       AJ851291     13079 bp    DNA     circular SYN 12-NOV-2004
DEFINITION  Cloning vector pRU1161 gusA gene for beta-glucuronidase and mRFP1
            gene for mono Red fluorescent protein.
ACCESSION   AJ851291
VERSION     AJ851291.1  GI:55724891
KEYWORDS    beta-glucuronidase; gusA gene; mono Red fluorescent protein; mRFP1
            gene.
SOURCE      Cloning vector pRU1161
  ORGANISM  Cloning vector pRU1161
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Karunakaran,R. and Poole,P.S.
  TITLE     High throughput promoter probe vectors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13079)
  AUTHORS   Poole,P.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2004) Poole P.S., School of AMS, University of
            Reading, Whiteknights, Reading, RG6 6AJ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..13079
                     /organism="Cloning vector pRU1161"
                     /mol_type="other DNA"
                     /db_xref="taxon:299186"
                     /lab_host="Escherichia coli"
     gene            186..863
                     /gene="mRFP1"
```

```
Qy      21 ThrValAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           :::|||||||||||||||||:|||||:||||||||||||||||||||||||||||||||||
Db     386 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 445
```

```
Qy      41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     446 ACCCAGACACCGCACCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 505
```

```
Qy      61 LeuSerProGlnPheGlnTyrGlySerLysAlaTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     506 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 565
```

```
Qy      81 AspTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     566 GACTACCTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 625
```

```
Qy     101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     626 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 685
```

```
Qy     121 LysValLysLeuArgGlyThrAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     686 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 745
```

```
Qy     141 MetGlyTrpGluAlaSerThrGluArgMetTyrProGluAspGlyAlaLeuLysGlyGlu 160
           |:||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db     746 ATGGGCTGGGAGGCCTCCACCGAGCGCATGTACCCCGAGGACGGCGCCCTGAAGGGCGAG 805
```

```
Qy     161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrAspAlaGluValLysThrTyrLeuVal 180
           ||||||||||||||||||:::
Db     806 ATCAAGATGAGGCTGAAGCTGAAAGACGGGCCACTACGACGCCGAGGTCAAGACCTAC 865
```

```
Qy     181 TyrMetAlaLysLysProValGlnLeuProGlyAlaTyrLysThrAspIleLysLeuAsp 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     866 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGAC 925
```

```
Qy     201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     926 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC 985
```

```
Qy     221 His 221
           |||
Db     986 CAC 988
```

ORIGIN

Alignment Scores:

Pred. No.:   1

Score: 8.78e-94

Percent Similarity: 89.14%

Best Local Similarity: 86.88%

Query Match: 84.10%

Length:	5311
Matches:	192
Conservative:	5
Mismatches:	24
Indels:	0
Gaps:	0

US-10-006-922a-12 (1-225) x AJ851287 (1-5311)

```
Qy       1 MetArgSerArgLysArgAsnValIleGluPheMetArgPheLysValArgMetGluGly 20
           |||||:||||||||||:|:||||||||||||||||||||||||||||||||||||||||
Db     326 ATGGCGCTCCGGAGGACGTCATCGAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 385
```

CDS   186..863
       /gene="mRFP1"
       /codon_start=1
       /transl_table=1
       /product="mono Red fluorescent protein"
       /db_xref="GI:55724892"
       /protein_id="CAH64921.1"
       /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
       KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
       GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
       IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKIDIKLDITSHNEDYTIVEQYERAE
       GRHSTGA"

gene   885..2696
        /gene="gusA"

CDS   885..2696
        /gene="gusA"
        /codon_start=1
        /transl_table=1
        /product="beta-glucuronidase"
        /protein_id="CAH64922.1"
        /db_xref="GI:55724893"
        /translation="MVRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRA
        IAVPGSFNDQFADADIRNYAGNVNYQREVEIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
        EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTIPPGMVITDENGKKKQSY
        FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
        ADQQVVATGQGTSGTLQVVNPHLMQPGEGYLYELCVTAKSQTEDIYPLRVGIRSVAV
        KGEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
        EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAI
        KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
        AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQERLHQPIIITEYGVD
        TLAGLHSMYTDMWSEEYQCAWLDMTHRVFDRVSAVVGEQVWNFADFATSQGILRVGGN
        KKGIFTRDRKPKSAAFLLQKRWTGMNFGRKPQQGGKQ"

US-10-006-922A-12 (1-225) x AJ851291 (1-13079)

ORIGIN

Alignment Scores:
Pred. No.:
Score: 2.64e-93
Percent Similarity: 1021.00
Best Local Similarity: 89.14%
Query Match: 86.88%
                     84.10%
DB: 11

Length: 13079
Matches: 192
Conservative: 5
Mismatches: 24
Indels: 0
Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        ||| |||||| |::|||::::|||::||:||||||||||||||||||||||||||||||||||
Db 186 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 245

Qy   21 ThrValAsnGlyHisGluPheIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        :::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 305

Qy   41 HisAsnThrValLysLysPheLysIleThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 306 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 365

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysAlaTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 366 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCC 425

Qy   81 AspTyrLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 426 GACTACCTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 485

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db 486 GACGGCGGCGTGGTGACCGTGACCCAGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 545

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db 546 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 605

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

US-10-006-922A-12 (1-225) x AJ851291 (1-13079)

Db 606 ATGGGCTGGGAGGCCTTCCACCGAGACGGCGATGTACCCCGAGGCACGGCGGCGGAGTGCGCCGCCCCTGAAGGGCGGAG 665
        |||||||||||||||||||||||||:|:||:||||||||||||||||||||||||||||||||||| |||||||||||||

Qy 161 IleHisLysAlaLeuLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||:|:|||||||||||||||:::|
Db 666 ATCAAGAAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACC 725

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 726 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTCGAC 785

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 786 ATCACCTCCCACAACGAGGACTACACCATCGTGCAGCAGTACGAGCGCGCCGAGGGGCGC 845

Qy 221 His 221
        |||
Db 846 CAC 848

LOCUS   AY678271
DEFINITION Synthetic construct monomeric yellow fluorescent protein gene,
            complete cds.
ACCESSION   AY678271
VERSION   AY678271.1   GI:55420626
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM synthetic construct
            other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 678)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
            Palmer,A.E. and Tsien,R.Y.
  TITLE   Improved monomeric red, orange and yellow fluorescent proteins
            derived from Discosoma sp. red fluorescent protein
  JOURNAL   Nat. Biotechnol. 22 (12), 1567-1572 (2004)
  PUBMED   15558047
REFERENCE   2 (bases 1 to 678)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
            Tsien,R.Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Pharmacology, University of California, San
            Diego, 9500 Gilman Dr., La Jolla, CA 92093-0647, USA
FEATURES   Location/Qualifiers
     source   1..678
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS   1..678
                     /codon_start=1
                     /transl_table=1
                     /note="mHoneydew; engineered variant of monomeric red
                     fluorescent protein mRFP1 in GenBank Accession Number
                     AF506027"
                     /product="monomeric yellow fluorescent protein"
                     /protein_id="AAV52171.1"
                     /db_xref="GI:55420627"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFMWGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWAATTERNYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKIDGKLDITSHNEDYTIVEQYERAE
                     GRHSTGA"

ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.92e-93
Percent Similarity: 1007.00
Best Local Similarity: 88.69%
Query Match: 85.07%
                     82.95%
DB: 11

Length: 678
Matches: 188
Conservative: 8
Mismatches: 25
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AY678271 (1-678)

RESULT 78

```
LOCUS       AY678270                678 bp    DNA     linear   SYN 17-DEC-2004
DEFINITION  Synthetic construct monomeric orange-red fluorescent protein gene,
            complete cds.
ACCESSION   AY678270
VERSION     AY678270.1  GI:55420624
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
            Palmer,A.E. and Tsien,R.Y.
  TITLE     Improved monomeric red, orange and yellow fluorescent proteins
            derived from Discosoma sp. red fluorescent protein
  JOURNAL   Nat. Biotechnol. 22 (12), 1567-1572 (2004)
  PUBMED    15558047
REFERENCE   2  (bases 1 to 678)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
            Tsien,R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Pharmacology, University of California, San
            Diego, 9500 Gilman Dr., La Jolla, CA 92093-0647, USA
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS             1..678
                     /note="mTangerine; engineered variant of monomeric red
                     fluorescent protein mRFP1 in GenBank Accession Number
                     AF506027"
                     /codon_start=1
                     /transl_table=11
                     /product="monomeric orange-red fluorescent protein"
                     /protein_id="AAV52170.1"
                     /db_xref="GI:55420625"
                     /translation="MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFCYGSKAYVKHPADIPDYLKLSPEGFKWERVMNFEDG
                     GVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVELYERAE
                     GRHSTGA"
ORIGIN
```

US-10-006-922A-12 (1-225) x AY678270 (1-678)

Alignment Scores:
```
Pred. No.:                            6.22e-93
Score:               1002.00          Length:        678
Percent Similarity:  88.24%           Matches:       189
Best Local Similarity: 85.52%         Conservative:    6
Query Match:         82.54%           Mismatches:     26
DB:                  11               Indels:          0
                                      Gaps:            0
```

```
Qy     1 MetArgSerSerLysAsnValIleLeuLysGluProPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         :::||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         :::||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCTCAGTTCATGTGGGGGTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGGTGCTGAGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||:::
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||:::
Db   481 ATCAAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGGTCAAGACCAC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         :::||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGGCGCGAGGGGCGC 660

Qy   221 His 221
         |||
Db   661 CAC 663
```

RESULT 79

```
Db   AY679163
LOCUS       AY679163                 702 bp    DNA     linear   SYN 21-NOV-2004
DEFINITION  Synthetic construct mRFPmars-HIS6 gene, complete cds.
ACCESSION   AY679163
VERSION     AY679163.1  GI:55793413
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Fischer,M., Haase,I., Simmeth,E., Gerisch,G. and
            Muller-Taubenberger,A.
  TITLE     A brilliant monomeric red fluorescent protein to visualize
            cytoskeleton dynamics in Dictyostelium
  JOURNAL   FEBS Lett. 577 (1-2), 227-232 (2004)
   PUBMED   15527790
REFERENCE   2  (bases 1 to 702)
  AUTHORS   Fischer,M.J. and Mueller-Taubenberger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2004) Chemistry, Technical University Munich,
            Lichtenbergstr. 4, Garching 85747, Germany
FEATURES             Location/Qualifiers
     source          1..702
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS             1..702
                     /note="red fluorescent protein"
                     /codon_start=1
                     /transl_table=11
                     /product="mRFPmars-HIS6"
                     /protein_id="AAV65051.1"
                     /db_xref="GI:55775541"
                     /translation="MASSEDVIKEPMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTA
                     KLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERIMNFEDG
                     GVVTVQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGE
                     IKMRLKLKDGGHYDAEVKTTYKAKKQVLPGAYKTDIKLDIISHNEDYTIVEQYERAE
                     GRHSTGAGSHHHHHH"
     misc_feature    676..681
                     /note="Region: linker"
     misc_feature    682..699
                     /note="Region: histidine-tag; 6xHis"
ORIGIN
```

US-10-006-922A-12 (1-225) x AY679163 (1-702)

```
Alignment Scores:
Pred. No.:
Score:                  994.00
Percent Similarity:     87.78%
Best Local Similarity:  84.16%
Query Match:            81.88%
DB:

Length:         702
Matches:        186
Conservative:     8
Mismatches:      27
Indels:           0
Gaps:             0

Score:         4.26e-92
```

```
Qy    1 MetArgSerSerLysAsnValIleLeuGlyPheMetArgPheLysValArgMetGluGlyGly 20
          ::|||||:||::::|||||||| 
Db    1 ATGGCATCRTCAGAAGACGTTATTAAAGAATTTATGAGATTCAAAGTTAGAATGGAAGGT 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
Db   61 TCAGTTAACGGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACCATATGAAGGT 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 ACAGAAACAGCAAAATTAAAAGTTACAAAAGGTGGTCCATTACCATTTGCATGGGATATT 180
```

RESULT 80

```
Db   AY678265
LOCUS       AY678265                 711 bp    DNA     linear   SYN 17-DEC-2004
DEFINITION  Synthetic construct monomeric orange fluorescent protein gene,
            complete cds.
ACCESSION   AY678265
VERSION     AY678265.1  GI:55420614
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
            Palmer,A.E. and Tsien,R.Y.
  TITLE     Improved monomeric red, orange and yellow fluorescent proteins
            derived from Discosoma sp. red fluorescent protein
  JOURNAL   Nat. Biotechnol. 22 (12), 1567-1572 (2004)
   PUBMED   15558047
REFERENCE   2  (bases 1 to 711)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
            Tsien,R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Pharmacology, University of California, San
            Diego, 9500 Gilman Dr., La Jolla, CA 92093-0647, USA
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS             1..711
                     /note="mOrange; engineered variant of monomeric red
                     fluorescent protein mRFP1 in GenBank Accession Number
                     AF506027; N- and C-terminal replaced with equivalent
                     residues from EGFP"
                     /codon_start=1
                     /transl_table=11
                     /product="monomeric orange fluorescent protein"
```

```
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysLysValTyrValLysHisProAlaAspIlePro 80
Db  181 CTTTCACCACAATTTCAATATGGTAGTAAAAGCATATGTTAAACATCCAGCAGATATTCCA 240

Qy   81 AspTyrLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GATTATCTTAAACTTTCATTTCCAGAAGGTTTTAAATGGGAAAGAGTTATGAATTTTGAA 300

Qy  101 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuGlyGlu 120
Db  301 GATGGGTGGTGTGTAACAGTTACTCAAGATTCAAGCATTACAAGAGGTGAATTTATTTAT 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAAGTTAAAATTAAGAGTGTACCAGTTAATGCAAAAAGGTGAAGGTGAAGGTAAAAAACT 420

Qy  141 MetGlyTrpGluAlaSerSerGluArgLeuTyrProGluAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGTTGGGGAAGCATCAACAGAGAAAGAAATGTATCCAGAAGACGGTGTTCTAAAAGGT 480

Qy  161 IleHisLysAlaLeuLysLeuLysProGlyGlyTyrTyrHisTyrLeuValLysHisSerIle 180
Db  481 ATTAAAAATGAGATTAAAAACTTAAAACAGAAAAGGTTATTAACAAAAACACTTAACAGAA 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaSerLysLeuAsp 200
Db  541 TATAAAAGCTAAAAACAAGTTCAAAACCAGGTGCTTATTATAAAAGATATTAAAATAGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnGlnTyrGluArgThrGluGlyArg 220
Db  601 ATTATTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGACGAGAAGGTAGA 660

Qy  221 His 221
Db  661 CAT 663
```

```
ORIGIN
                    /protein_id="AAV52165.1"
                    /db_xref="GI:55420615"
                    /translation="MVSKGEENNMAIIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYE
                    GTQTAKLKVTKGGPLPFAWDILSPQFTYGSKAYVKHPADIPDYFKLSFPEGFKWERVM
                    NFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDG
                    ALKGEIKMRLKLKDGGHYTSEVKTTYKAKKPVQLPGAYIVGIKLDITSHNEDYTIVEQ
                    YERAEGRHSTGGMDELYK"

US-10-006-922A-12 (1-225) x AY678265 (1-711)

Qy    7   ValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlyGlu  26
          ||||||||||||||||||||||||||||||||||| |:|||||| ||||||||||||||||||
Db    34  AtcAtcAagGagGagTtcAtgCgCttCaaGgtGcgGatGgaGggCtcCgtGaAcGgcCacGag  93

Qy    27  PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyThrGlnThrAlaLysLeu  46
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    94  TtcGagAtcGagGggGagGggGagGggCgcCccTacGagGggCttCagactGccTagAgcTg  153
```

ORIGIN

Qy 47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66

(Sequence alignment data — Qy/Db nucleotide alignment continues)

RESULT 82
AY786536

LOCUS   AY786536   681 bp   mRNA   linear   INV 01-DEC-2004
DEFINITION   Discosoma sp. LW-2004 fluorescent protein raspberry mRNA, complete cds.
ACCESSION   AY786536
VERSION   AY786536.1 GI:55792806
KEYWORDS   .
SOURCE   Discosoma sp. LW-2004
  ORGANISM   Discosoma sp. LW-2004
             Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
             Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1 (bases 1 to 681)
  AUTHORS   Wang,L., Jackson,W.C., Steinbach,P.A. and Tsien,R.Y.
  TITLE   Evolution of new nonantibody proteins via iterative somatic
             hypermutation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (48), 16745-16749 (2004)
   PUBMED   15556995
REFERENCE   2 (bases 1 to 681)
  AUTHORS   Wang,L., Jackson,W.C., Steinbach,P.A. and Tsien,R.Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-OCT-2004) Department of Pharmacology, University of
             California-San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES   Location/Qualifiers
     source   1..681
                     /organism="Discosoma sp. LW-2004"
                     /mol_type="mRNA"
                     /db_xref="taxon:301246"
     CDS   1..681
                     /codon_start=1
                     /product="fluorescent protein raspberry"
                     /protein_id="AAV65486.1"
                     /db_xref="GI:55792807"
                     /translation="MVSKGEEVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQT
                     AKLKVTKGGPLPFAWDILSPQCMYGSKGYVKHPADIPDYLKLSFPEGFKWERVMNFED
                     GGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKG
                     EMKMRLKLKDGGHYDAEVKTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERA
                     EGRHSTGA"

ORIGIN

Qy 147 ThrGluArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlyGluLeuHisLysAlaLeuLys 166
Db 454 TCCGAGCGGGATGTACCCGCGGGACGGCGCCCTGAAGGGCGAGATGAAGATGAGGCTGAAG 513
Qy 167 LeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleLeuTyrMetAlaLysLysPro 186
Db 514 CTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCTACATGGCCAAGAAGCCC 573
Qy 187 ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu 206
Db 574 GTGCAGCTGCCCGGGCCCTTACAACGTCATCAAGACTTGGACATCACCTCCCACAACGAG 633
Qy 207 AspThrTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221
Db 634 GACTACACCATCGTTGAACAGTACGAACGCCCCGAGGGCCGCCAC 678

RESULT 83
AY678266

LOCUS   AY678266   711 bp   DNA   linear   SYN 17-DEC-2004
DEFINITION   Synthetic construct monomeric red-orange fluorescent protein gene,
             complete cds.
ACCESSION   AY678266
VERSION   AY678266.1 GI:55420616
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 711)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
             Palmer,A.E. and Tsien,R.Y.
  TITLE   Improved monomeric red, orange and yellow fluorescent proteins
             derived from Discosoma sp. red fluorescent protein
  JOURNAL   Nat. Biotechnol. 22 (12), 1567-1572 (2004)
   PUBMED   15558047
REFERENCE   2 (bases 1 to 711)
  AUTHORS   Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
             Tsien,R.Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Pharmacology, University of California, San
             Diego, 9500 Gilman Dr., La Jolla, CA 92093-0647, USA
FEATURES   Location/Qualifiers
     source   1..711
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS   1..711
                     /codon_start=1
                     /note="mStrawberry; engineered variant of monomeric red
                     fluorescent protein mRFP1 in GenBank Accession Number
                     AF506027; N- and C-termini replaced with equivalent
                     residues from EGFP"

Db   76 CACGAGGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCC 135
Qy   45 LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln 64
Db 136 AAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAG 195
Qy   65 PheGlnTyrGlySerIleTyrValValLysHisAlaProAlaAspIleProAspTyrLysLys 84
Db 196 TGCATGTACGGGCTCCAAGGGCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAG 255
Qy   85 LeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyVal 104
Db 256 CTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAGGACGGCGGCGTG 315
Qy 105 ValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyrLysValLysPhe 124
Db 316 GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGTTC 375
Qy 125 IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu 144
Db 376 CGCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG 435
Qy 145 AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlyGluHisLysAla 164
Db 436 GCCTCCACCGAGCGGCTGTACCCCGACGGCGGCGCCCTGAAGGGCGGCGAGATGAAGATGAGG 495
Qy 165 LeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleLeuTyrMetAlaLys 184
Db 496 CTGAAGCTGAAGGACGGCGGCCACTACCTGGTCGAACTGAAGAGCATCCTGTACATGGCCAAG 555
Qy 185 LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis 204
Db 556 AAGCCCGTGCAGCTGCCCGGGCCCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCAC 615
Qy 205 AsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221
Db 616 AACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGCCAC 666

US-10-006-922A-12 (1-225) x AY786536 (1-681)

Alignment Scores:
Pred. No.:
Score: 3.41e-91
                      995.00
Percent Similarity: 88.02%
Best Local Similarity: 85.25%
Query Match: 81.14%

Length: 681
Matches: 185
Conservative: 6
Mismatches: 26
Indels: 0
Gaps: 0
DB: 2

Qy   5 LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly 24
Db   16 GAGGAGGTCATCAAGGAGTTCATGCGGCTTCAAGGTGCCGCATGGAGGGCGTTCAGCACC 75
Qy   25 HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal 44

Left column

```
ORIGIN
        /transl_table=11
        /product="monomeric red-orange fluorescent protein"
        /protein_id="AAV52166.1"
        /db_xref="GI:55420617"
        /translation="MVSKGEENNMAIIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYE
        GTQTAKLKVTKGGPLPFAWDILTPNFTYGSKAYVKHPADIPDYLKLSFPEGFKWERVM
        NFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDG
        ALKGEIKMRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYIVGIKLDITSHNEDYTIVEL
        YERAEGRHSTGGMDELYK"
```

US-10-006-922A-12 (1-225) x AY678266 (1-711)

Alignment Scores:
Pred. No.:
Score: 6.03e-90
                                973.00
Percent Similarity: 87.44%
Best Local Similarity: 84.65%
Query Match: 80.15%
DB: 11

Length: 711
Matches: 182
Conservative: 6
Mismatches: 27
Indels: 0
Gaps: 0

```
Qy    7  ValIleLysGlyGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu   26
         :::|||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db   34  AtCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGGCACGAG      93

Qy   27  PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyThrGlnHisAsnThrValLysLeu   46
         |||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db   94  TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTG      153

Qy   47  LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln   66
         ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db  154  AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTAAGCCCCCAACTTCACC   213

Qy   67  TyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLeuLysLeuSer   86
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  214  TACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCC   273

Qy   87  PheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValValThr  106
         :::||||||||||||||||||||||||||||||||||||||||||:::|||||||||||
Db  274  TTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGGGGCGTGGTGACC  333

Qy  107  ValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIleGly  126
         ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db  334  GTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTCAAGTTCATCGGC  393

Qy  127  ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer  146
         ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db  394  ACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAAGCCTCC  453

Qy  147  ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysValAlaLeuLys  166
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  454  TCCGAGCGGGAGTACCCCGAGGACGGCGTCCTGAAGGGCGAGATCAAGATGAGGCTGAAG  513

Qy  167  LeuLysAspGlyValHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro  186
         ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db  514  CTGAAGGACGGCGGCCACTACGACGTGGAGTTCAAGTCAATCTACACGGCCAAGAAGCCC  573

Qy  187  ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu  206
         |||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db  574  GTGCAGCTGCCCGGCGCCTACTACTATGTGGACATCAAGTTGGACATCACCTCCCACAACGAG  633

Qy  207  AspTyrThrIleValGluLeuGlnThrArgIleThrGlyLeuArgHis  221
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  634  GACTACACCATCGTGGAACTGCTACGAACGGCGCCCGAGGGCCGCCAC  678
```

RESULT 84

Right column

ORIGIN

US-10-006-922A-12 (1-225) x AY786537 (1-681)

Alignment Scores:
Pred. No.:
Score: 7.24e-90
                                972.00
Percent Similarity: 87.10%
Best Local Similarity: 83.41%
Query Match: 80.07%
DB: 2

Length: 681
Matches: 181
Conservative: 8
Mismatches: 28
Indels: 0
Gaps: 0

```
Qy    5  LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly   24
         :::||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db   16  GAGGAGGTCATCAAGGAGTTCATGCGCTTCAAGGAGCACATGGAGGGCTCCGTGAACGGC   75

Qy   25  HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyThrGlnThrAla   44
         |||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db   76  CACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCC  135

Qy   45  LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln   64
         :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136  AGGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAG  195

Qy   65  PheGlnIleTyrGlySerLysAlaTyrValLysHisProAlaAspIleProAspTyrLys   84
         ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db  196  ATCATGTACGGGTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAG  255

Qy   85  LeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyVal  104
         |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||
Db  256  CTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTG  315

Qy  105  ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe  124
         |||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db  316  GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTC  375

Qy  125  IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu  144
         ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db  376  CGCGGGGACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG  435

Qy  145  AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAla  164
         ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db  436  GCCTCCTCCGACGGGAGTACCCCGAGGACGGCGTCCTGAAGGGCGAGATCAAGATGAGG  495
```

```
RESULT 85
AY678267

LOCUS           AY678267                 711 bp    DNA     linear   SYN 17-DEC-2004
DEFINITION      Synthetic construct monomeric yellow-orange fluorescent protein
                gene, complete cds.
ACCESSION       AY678267
VERSION         AY678267.1  GI:55420620
KEYWORDS        .
SOURCE          synthetic construct
  ORGANISM      synthetic construct
                other sequences; artificial sequences.
REFERENCE       1  (bases 1 to 711)
  AUTHORS       Shaner,N.C., Campbell,R.E., Steinbach,P.A., Giepmans,B.N.,
                Palmer,A.E. and Tsien,R.Y.
  TITLE         Improved monomeric red, orange and yellow fluorescent proteins
                derived from Discosoma sp. red fluorescent protein
  JOURNAL       Nat. Biotechnol. 22 (12), 1567-1572 (2004)
  PUBMED        15558047
REFERENCE       2  (bases 1 to 711)
  AUTHORS       Shaner,N.C., Campbell,R.E., Steinbach,P.A., Palmer,A.E. and
                Tsien,R.Y.
  TITLE         Direct Submission
  JOURNAL       Submitted (06-JUL-2004) Pharmacology, University of California, San
                Diego, 9500 Gilman Dr., La Jolla, CA 92093-0647, USA
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS             1..711
                     /note="mBanana; engineered variant of monomeric red
                     fluorescent protein mRFP1 in GenBank Accession Number
                     AF506027; N- and C-termini replaced with equivalent
                     residues from EGFP"
                     /codon_start=1
                     /transl_table=11
                     /product="monomeric yellow-orange fluorescent protein"
                     /protein_id="AAV52167.1"
                     /db_xref="GI:55420621"
                     /translation="MVSKGEENNMAVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYE
                     GTQTAKLKVTKGGPLPFAWDILSPQFCYGSKAYVKHPTGIPDYFKLSFPEGFKWERVM
                     NFEDGGVVTVAQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDG
                     ALKGEIKMRLKLKDGGHYSAETKTTYKAKKPVQLPGAYIAGEKIDITSHNEDYTIVEL
                     YERAEGRHSTGGMDELYK"
ORIGIN
```

Alignment Scores:

Pred. No.:

Score:	960.00
Percent Similarity:	86.05%
Best Local Similarity:	83.72%
Query Match:	79.08%

DB:

```
US-10-006-922A-12 (1-225) x AY678267 (1-711)
```

Length:	711	
Matches:	180	
Conservative:	5	
Mismatches:	30	
Indels:	0	
Gaps:	0	

```
Qy      7 ValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu  26
          ||||||||||||||||||||||||||||||||||||||:::||||||||::||||||||||
Db     34 GTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAG  93

Qy     27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValValLeu  46
          |||||||||||||||||||||||||||||||||||||||||
Db     94 TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCCCACCGTCGTGCTG 153

Qy    165 LeuLysLeuLysAspIleArgGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys 184
          ||||||||::::||||||||||||||||||||||||||||||||||||::|||||||||||||
Db    496 CTGAGGCTGAAGGACGGCGGCCACTACAGCGCCGAGACCAAGACCACCTACAAGGCCAAGCTG 555

Qy    185 LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis 204
          |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db    556 AAGCCGGTGCAGCTGCCCGGCGCCTACAAGCACCGACATCAAGCTGGACATCACCTCCCAC 615

Qy    205 AsnGluAspTyrThrIleValGluLeuGlnGlnTyrGlyArgThrGluGlyArgHis 221
          |||||||||||||||||||||||||||||||||||||||||
Db    616 AACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCAC 666

Qy     67 TyrGlySerLysTyrValTyrValLysHisProAlaAspTyrLysLysLeuSer 86
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    214 TACGGCTCCAAGGCCTACGTGAAGCACCCCGCAGACTACAAGAAGCTGTCC 273

Qy    214 AspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221
          |||||||||||||||||||||||||||||||||||||||||||||
Db    634 GACTACACCATCGTGGAATTGTACGAGCGCACCGAGGGCCGCCAC 678

Qy    207 AspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    574 GTGCAGTTGCCCGGCGCCTACAAGCACCGGCGAGAAGATCGACATCACCTCCCACAATGAG 633

Qy    187 ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu 206
          ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db    514 CTGAAGGACGGCGGCCACTACAGCGCCGAGACCAAGACCACCTACAAGGCCAAGAAGCCC 573

Qy    167 LeuLysAspIleArgGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    454 TCCGAGCGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAAG 513

Qy    147 ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleLysHisLysAlaLeuLys 166
          :::||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    394 ACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC 453

Qy    127 ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer 146
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    334 GTGGCTCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGC 393

Qy    107 ValThrValAlaGlnAspSerSerLeuGlnAspGlyGluPheIleTyrLysValLysLeuArg 126
          |||||||||||||||||||||||||||||||::|||||||||||||||||||||||||||
Db    274 TTCCCCGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333

Qy     87 PheProGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValValThr 106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    107 GTGGGCTCAGGACTTCCTCCCTGCAGGACGGCGGAGTTCATCTACAAGGTGAAGCTG 166

Qy     47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
          |||:|||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db    154 AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCGTGGGACATCCTGTCCCCTCAGTTCTGC 213

Qy     94 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCGTGGGACATCCTGTCCCCTCAGTTCTGC 66
```

Search completed: January 12, 2006, 13:57:52
Job time : 3773 secs

```
                      GenCore version 5.1.6
                      Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         January 12, 2006, 11:20:58 ; Search time 478 Seconds
                                              (without alignments)
                                              3137.146 Million cell updates/sec

Title:          US-10-006-922A-12
Perfect score:  1214
Sequence:       1 MRSSKNVIKPMRFKVRMEG............EDYIVEQYERTEGRHHLFL 225

Scoring table:  BLOSUM62
                Xgapop  10.0 ,  Xgapext  0.5
                Ygapop  10.0 ,  Ygapext  0.5
                Fgapop   6.0 ,  Fgapext  7.0
                Delop    6.0 ,  Delext   7.0

Searched:       4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 83
                Listing first 500 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 69%
                 Maximum Match 100%

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat_10012006_162503_19106/app_query.fasta_1.391
-DB=N_Geneseq -QFMT=fastap -SUFFIX=p2nminscr.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=500 -DOCALIGN=200 -NORM=ext -THR_MAX=100 -THR_MIN=69 -ALIGN=500
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10006922 @CGN_1_1.1096 @runat_10012006_162503_19106 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
       1:  N_Geneseq_21:*
       2:  geneseqm1980e:*
       3:  geneseqm1990e:*
       4:  geneseqm2000e:*
       5:  geneseqm2001aa:*
       6:  geneseqm2001ba:*
       7:  geneseqm2002aa:*
       8:  geneseqm2002ba:*
       9:  geneseqm2003aa:*
      10:  geneseqm2003ba:*
      11:  geneseqm2003ca:*
      12:  geneseqm2003da:*
      13:  geneseqm2004aa:*
      14:  geneseqm2004ba:*
           geneseqm2005s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Match Length	DB	ID	Description
1	1214	100.0	678	4	AAH47654	Discosoma
2	1214	100.0	678	4	AAD13053	Discosoma
3	1214	100.0	678	6	AAD11142	Discosoma
4	1214	100.0	678	6	ABA95905	Yeast opt
5	1214	100.0	678	6	AAA28207	Discosoma
6	1214	100.0	678	6	ABA95921	Yeast opt
7	1214	100.0	678	6	ABA95920	Yeast opt
8	1214	100.0	695	3	ADM97768	D sp red
9	1214	100.0	859	3	AAA48743	Humanised
10	1214	100.0	678	10	ADC24127	Discosoma
11	1214	100.0	678	10	ADF70404	Discosoma
12	1214	100.0	678	11	ADL46204	Discosoma
13	1214	100.0	678	11	ADN33978	Wild-type
14	1214	100.0	678	12	ADI36420	Discosoma
15	1214	100.0	678	12	AAA48749	Humanised
16	1214	100.0	681	10	ADL46205	Human cod
17	1214	100.0	681	11	ADL46225	Human cod
18	1214	100.0	859	8	AAD53432	Discosoma
19	1214	100.0	859	7	AAD61969	Discosoma
20	1214	100.0	859	10	ADX26533	Discosoma
21	1210	99.7	859	14	ACA62995	Plasmid D
22	1210	99.7	3311	9	ADC24134	Discosoma
23	1210	99.7	681	10	ADC24134	Discosoma
24	1210	99.7	681	11	ADL46205	Human cod
25	1210	99.7	681	11	ADL46225	Human cod
26	1210	99.7	723	8	ABZ222476	Mammalian
27	1210	99.7	1638	10	ADL18131	RFP:PS(NI
28	1210	99.7	1647	10	ADL18155	RFP:PS(HI
29	1210	99.7	2721	13	ADS75466	Fibrohexa
30	1210	99.7	2772	13	ADS75468	Fibrohexa
31	1210	99.7	4692	6	AAL47954	Modified
32	1210	99.7	4692	10	ACC44640	Vector pD
33	1210	99.3	5436	4	AAD10003	Plasmid p
34	1210	99.7	6893	9	ADE24111	Proviral
35	1210	99.7	6985	10	AAD09979	pBiT(dHSP
36	1210	99.7	7910	4	ABS56664	Plasmid P
37	1210	99.7	9320	6	ADI04104	Vector pT
38	1210	99.7	9658	12	ADI04103	Vector pT
39	1210	99.7	9678	12	ADI04102	Vector pT
40	1210	99.6	10263	12	ABAS95906	Yeast opt
41	1209	99.6	681	6	ABAS95907	Yeast opt
42	1209	99.6	681	6	AAD28208	Discosoma
43	1207	99.3	678	4	AAH47656	Anthozoan
44	1205	99.1	681	13	ADR30814	Zebrafish
45	1203	98.6	6985	10	ADR30815	Zebrafish
46	1196	98.5	666	9	ABL61142	Red fluor
47	1196	98.5	711	12	ACA55359	Transform
48	1196	98.5	711	12	ADO78067	Corallimo
49	1196	98.5	711	14	ADO78069	Corallimo
50	1196	98.5	711	14	ADZ84224	Actinodis
51	1196	98.1	678	6	ADZ84222	Actinodis
52	1191	98.1	678	6	AAD46282	Discosoma
53	1191	97.7	675	6	AAD28209	Discosoma
54	1186	97.7	747	14	AAD46281	Discosoma
55	1186	97.7	4200	14	ADV96699	Plant cel
56	1186	97.7	4300	14	ADW43589	UbCP-lox-
57	1186	97.7	4488	13	ADW43588	UbCP-lox-
58	1186	97.7	6706	12	ADQ91220	Fusion co
59	1186	97.7	7495	13	ADQ91222	Plasmid p
60	1186	97.7	7616	12	ADR30815	Zebrafish
61	1186	97.7	7927	12	ADL35208	Plasmid D
62	1186	97.7	9731	11	ADL35211	Plasmid p
63	1186	97.7	9782	11	ADM82791	DNA repai
64	1184	97.5	898	11	ADM82792	DNA repai
65	1184	97.5	898	4	AAA48742	Discosoma
66	1160	95.6	678	4	AAD03615	Discosoma
67	1160	95.6	678	11	ADC24128	Discosoma
68	1160	95.6	704	11	ADL46207	Discosoma
69	1132	93.2	678	11	ADN33980	Discosoma
70	1121	92.3	681	12	ADM13536	DNA encod
71	1121	92.3	681	11	ADC24130	Discosoma
72	1119	92.3	678	11	ADL46209	Discosoma
73	1119	92.2	678	6	AAA48748	Humanised
74	1111	91.5	678	6	AAD46280	Discosoma
75	1105	91.0	678	3	AAA48749	Humanised
76	1085.5	89.4	876	3	AAA48750	Humanised
77	1085.5	89.4	876	4	AAA28958	Discosoma
					AAA03618	Discosoma

```
US-10-006-922A-12 (1-225) x AAH47654 (1-678)

Query Match:          100.00%              Gaps:         0
Best Local Similarity: 100.00%             Indels:       0
Percent Similarity:    100.00%             Mismatches:   0
Percent Identity:      100.00%             Conservative: 0
Score:                 1214.00             Matches:      225
Pred. No.:                                 Length:       678
                       2.3e-140

Alignment Scores:
```

SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

XX
CC   anthozoan fluorescent protein
CC   represents the coding sequence of a Discosoma sp. drFP583 protein, an
CC   emission peaks when expressed in mammalian cells. The present sequence
CC   brightness, and create (II) with sharper, more defined excitation and
CC   The key mutations in the encoding nucleic acids provide improved folding,
CC   cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC   fluorescent proteins. (II) has improved brightness, reduced spectral
CC   fluorescence analysis and FRET-based applications using existing spectral
CC   protein-protein interactions. (II) is also suitable for multiplexed
CC   invention can be expressed by standard recombinant methodology. (II) are
CC   used a fluorescent markers and FRET partners. It is used for identifying
CC   different fluorescent properties. The red fluorescent protein, of the
CC   functional red fluorescent proteins, and the encoding nucleic acids, with
CC   protein by at least one amino acid substitution, and with
CC   fluorescent protein (II) that differs from the sequence of an Anthozoan red
CC   The invention provides a nucleic acid encoding functional red fluorescent
XX
PS   Disclosure; Page 83; 90pp; English.
XX
PR   23-FEB-2000; 2000US-0184732P.
PF   13-FEB-2001; 2001WO-US004625.
PD   30-AUG-2001.
PN   WO20016291; 2001WO-US004625.
PT   Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
PT   key mutations for improving the proteins function.
PT   functional red fluorescent proteins, and the encoding nucleic acids, with
XX
PI   Nelson D, Zamaira E, Tsien R;
XX
PA   (AURO-) AURORA BIOSCIENCES CORP.
XX
OS   Discosoma sp.
XX
KW   anthozoan fluorescent protein
KW   Fluorescent protein; Anthozoan; fluorescence; marker; FRET; drFP583; 88.
XX
DE   Discosoma sp. red drFP583 protein coding sequence.
XX
DT   30-NOV-2001 (first entry)
XX
AC   AAH47654;
XX
ID   AAH47654 standard; cDNA; 678 BP.
XX
RESULT 1
AAH47654
```
      78  1085.5  89.4  876  10  ACC44649  Acc44649 Red fluor
      79  1021    84.4  678  10  ADC24132  Adc24132 Discosoma
      80  1021    84.1  678  11  ADI46211  Adi46211 Discosoma
      81  1021    84.1  2241 14  AEA54911  Aea54911 Fluoresce
      82  1015    83.6  675  12  ADQ59565  Adq59565 Discosoma
      83  1012    83.4  678  11  ADI46282  Adi46282 Discosoma
```

ALIGNMENTS

```
Qy      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAAGGTTCGGATGGAAGGA  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAAGGTTCGGATGGAAGGA  60

Qy     61 ACGGTCAATGGGCACCAGAGTTTGAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGC 120
```

RESULT 2
AAD13053

ID   AAD13053 standard; cDNA; 678 BP.
AC   AAD13053;
DT   16-OCT-2001 (first entry)
DE   Discosoma sp. "red" anthozoa fluorescent protein, drFP583 cDNA.
KW   Protein destabilisation; linker moiety; reporter moiety; disease
KW   transgenic animal; transgenic plant; disease resistance;
KW   linear multimerised domain; -NH-ubiquitin protein endoprotease;
DE   Protein destabilisation; linker moiety; reporter moiety; disease model;
OS   Discosoma sp.
PN   WO200157242-A2.
PD   09-AUG-2001.
PF   02-FEB-2001; 2001WO-US003791.
PR   04-FEB-2000; 2000US-00498098.
PA   (AURO-) AURORA BIOSCIENCES CORP.

```
XX
PI   Stack JH, Whitney M, Cubitt AB, Pollok BA;
XX
DR   WPI; 2001-488890/53.
XX
PT   Destabilizing proteins in living cells, by coupling a target protein to
PT   linear multimerized destabilization domain non-cleavable by -NH-ubiquitin
PT   protein endoproteases, comprising two copies of the domain.
XX
PS   Disclosure; Page 110; 171pp; English.
XX
CC   The present invention relates to a method for destabilising a target
CC   protein in a cell. The method comprises a linker moiety which operatively
CC   couples a target protein (a reporter moiety) to a linear multimerised
CC   destabilising domain, which is non-cleavable by a -NH-ubiquitin protein
CC   endoproteases. The method is useful for detecting an activity such as
CC   protease, protein kinase or phosphoprotein phosphatase activity and is
CC   also useful for identifying modulators of these activities. The method is
CC   also useful for developing novel assays for a wide range of post-
CC   translational activities, such as proteolysis, phosphorylation,
CC   dephosphorylation, glycosylation, methylation, sulfation, prenylation,
CC   disulfide bond formation and ADP-ribosylation within cells. The
CC   recombinant DNA molecule of the invention is useful for creating
CC   transgenic animals useful as disease models and transgenic plants with
CC   improved disease resistance or other favourable traits. The present
CC   sequence is Discosoma sp. "red" anthozoa fluorescent protein, dsRP483
CC   cDNA which is a natural fluorescent protein used as a reporter moiety in
CC   the exemplification of the invention
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AAD13053 (1-678)

Alignment Scores:
Pred. No.:              2.3e-140      Length:         678
Score:                  1214.00       Matches:        225
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                                   Gaps:           0

Qy    1 MetArgSerSerlysAsnValIleLysGlyGluPheMetArgPheLysValAlaArgMetGlyGly 20
Db    1 ATGAGGGTCTTCCAAGAATGTTATCAAGGAGAGTTCAAGGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGCCCATACGAAGGC 120

Qy   41 HisAsnThrValIleLysLeuValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTAAAGGTAAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerlysValThrTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCACAATTTCAGTTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrTyrLysLysLeuSerPheProGluGlyTyrLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGACGGACCCGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTGTTGAAAGGGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerlysLeuAsp 200
Db  541 TACATGGCCAAGAAGCCTGTGCAGCTACCAGGGTACTATTACGTTGACTCCAAACTTGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

```
RESULT 3
AAD11142
ID   AAD11142 standard; DNA; 678 BP.
XX
AC   AAD11142;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Discosoma species Anthozoa fluorescent protein, drFP583 encoding DNA.
XX
KW   Transmembrane potential; biological membrane; fluorescent ion; detection;
KW   test chemical screening; Anthozoa fluorescent protein; FP;
KW   transgenic organism; drFP583 protein; ds.
XX
OS   Discosoma sp.
XX
PN   WO200142211-A2.
XX
PI   Tsien RY, Gonzalez JE;
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PR   13-DEC-1999;   99US-00459956.
XX
PD   12-DEC-2000; 2000WO-US033739.
XX
PD   14-JUN-2001.
XX
DR   WPI; 2001-457276/49.
XX
PS   Disclosure; Page 150; 154pp; English.
XX
PT   Determining electrical potential across a membrane in biological systems,
PT   comprises introducing two reagents, exposing the membrane to light and
PT   measuring the energy transfer.
XX
CC   The patent discloses optical methods and compositions for determining
CC   transmembrane potential across biological membranes in living cells. The
CC   method of determining the electrical potential across a membrane in a
CC   biological system comprises introduction of two reagents, a first reagent
CC   comprising a hydrophobic fluorescent ion capable of redistributing from a
CC   first face of the membrane to a second face of the membrane in response
CC   to membrane potential change and a second reagent that label the first
CC   face or the second face of the membrane, which comprises a chromophore
CC   capable of undergoing energy transfer by either donating or accepting
CC   excited state energy to the fluorescent ion. The membrane is then exposed
CC   to excitation light and the energy transfer between the reagents is
CC   measured and related to the membrane potential. The method is useful for
CC   detecting changes in membrane potential in subcellular organelle
CC   membranes in biological systems. The method is used for screening of test
CC   chemicals for activity to modulate the activity of target ion channel.
CC   The invention also provides a transgenic organism comprising a first
CC   reagent that comprises a charged hydrophobic fluorescent molecule and a
CC   second reagent comprising a bioluminescent or naturally fluorescent
CC   protein. The present sequence is Discosoma species (red) DNA encoding an
CC   Anthozoa fluorescent protein (FP), drFP583
XX
```

SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.3e-140
Percent Similarity: 1214.00   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 100.00%   Conservative: 0
DB: 100.00%   Mismatches: 0
                           4   Indels: 0
                                          Gaps: 0

US-10-006-922A-12 (1-225) x AAD11142 (1-678)

Qy   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGGGAGGGCCATACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTCGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCCTT 675

XX
OS   Anthozoa.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /*tags= a
FT   /product= "yeast optimised Red Fluorescent Protein"
XX
PN   DE20001395-U1.
XX
DE   15-MAR-2001.
XX
PD   27-JAN-2000; 2000DE-02001395.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
DR   WPI; 2002-228394/29.
DR   P-PSDB; ABB08834.
XX
PT   New DNA encoding red fluorescent protein, useful as marker in
PT   biotechnology, has sequence optimised for expression in eukaryotes,
PT   especially yeast or plants.
XX
PS   Claim 1; Fig 1; 19pp; German.
XX
CC   The invention relates to DNA (I) containing either sequence ABA95905 or
CC   sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC   plants, especially dicotyledonous plants including Nicotiana tabacum or
CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern
CC   biotechnology. (I) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio
XX
SQ   Sequence 678 BP; 198 A; 147 C; 159 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.3e-140
Percent Similarity: 1214.00   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 100.00%   Conservative: 0
DB: 100.00%   Mismatches: 0
                           6   Indels: 0
                                          Gaps: 0

US-10-006-922A-12 (1-225) x ABA95905 (1-678)

Qy   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGAGATCTTCTAAGAACGTCATCAAGGAATTCATGAGACTTAAAGTTAGAATGGAAGGT 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACTGTTAACGGTCACGAATTCGAAATCGAAGGTGAAGGTGAAGGTAGACCATACGAAGGT 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAACACTGTCAAGTTGAAGGTTACTAAGGGTGGTCCATTGCCATTCGCTTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCTCCACAATTCCAATACGGTTCTAAGGTCTACGTCAAGCACCCAGCTGACATCCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAAGTTGTCCTTCCCAGAAGGTTTCAAGTGGGAAAGAGTCATGAACTTCGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 ...

RESULT 4
ABA95905

ID   ABA95905 standard; cDNA; 678 BP.
XX
AC   ABA95905;
XX
DT   29-MAY-2002 (first entry)
XX
DE   Yeast optimised RFP encoding cDNA SEQ ID NO 1.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology; gene; ss.

CC   biotechnology. (I) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio. The present
CC   sequence is that of a polynucleotide encoding the yeast optimised RFP,
CC   useful to the invention
XX
SQ   Sequence 678 BP; 202 A; 118 C; 159 G; 199 T; 0 U; 0 Other;

Alignment Scores:
```
Pred. No.:                         2.3e-140
Score:                             1214.00
Percent Similarity:                100.00%         Length:          678
Best Local Similarity:             100.00%         Matches:         225
Query Match:                       6               Conservative:    0
DB:                                                Mismatches:      0
                                                   Indels:          0
                                                   Gaps:            0
```

US-10-006-922A-12 (1-225) x ABA95922 (1-678)

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGATCTTCAAGAACGTTATTAAGGAGTTACTAAGGGTGGTCCATTGCCTTGGGACATT 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ...ACCGTTAACGGTCACGAATTCGAAATTGAAGGTGAAGGTGAAGGTAGACCATACGAAGGT 120

Qy   41 HisAsnThrValThrLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAACACTGTTAAGTTGAAGTGGTTACTAAGGGTGGTCCATTGCCTTGGGACATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTGTCTCCACAATTCCAATACGGTTCTAAGGTTTACGTTAAGCACCCAGCTGACATTCCA 240

Qy   81 AspTyrLysLysLeuSerPhePheGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTACAAGAAGTTGTCTTTCCCAGAAGGTTTCAAGTGGGAAAGAGTTATGAACTTCGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGTGTTGTTACTGTTACTCAAGACTCTTCTTTGCAAGACGGTTGTTTCATTTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTTAAGTTCATTGGTGTTAACTTCCCATCTGACGGTCCAGTTATGCAAAAGAAGACT 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGTTGGGAAGCTTCTACTGAAAGATTGTACCCAAGAGACGGTGTTTTGAAGGGTGAA 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATTCACAAGGCTTTGAAGTTGAAGGACGGTGGTCACTACTTGGTTGAATTCAAGTCTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCTAAGAAGCCAGTTCAATTGCCAGGTTACTACTACGTTGACTCTAAGTTGGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATTACTTCTCACAACGAAGACTACACTATTGTTGAACAATACGAAAGAACTGAAGGTAGA 660

Qy  221 HisHisHisLeuPheLeu 225
        ||||||||||||||||||
Db  661 CACCACTTGTTCTTG 675
```

XX
RESULT 6
ID   ABA95921 standard; DNA; 678 BP.
XX
AC   ABA95921;
XX
DT   29-MAY-2002 (first entry)

CC   biotechnology. (I) containing either sequence ABA95905 or
CC   sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC   plants, especially dicotyledonous plants including Nicotiana tabacum or
CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern

XX
DT   29-JAN-2000; 2000DE-02001395.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
DR   WPI; 2002-228394/29.
XX
PT   New DNA encoding red fluorescent protein, useful as marker in
PT   biotechnology, has sequence optimised for expression in eukaryotes,
PT   especially yeast or plants.
XX
PN   DE20001395-U1.
XX
PD   15-MAR-2001.

RESULT 5
ID   ABA95922 standard; DNA; 678 BP.
XX
AC   ABA95922;
XX
DT   29-MAY-2002 (first entry)
XX
DE   Yeast optimised RFP related DNA SEQ ID NO 19.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology; gene; ss.
OS   Anthozoa.
FH   Key   Location/Qualifiers
FH
FT   CDS   1..678
FT   /*tag=   a
FT   /product= "yeast optimised Red Fluorescent Protein"

```
Qy  221 HisHisHisLeuPheLeu 225
        ||||||||||||||||||
Db  661 CACCACTTGTTCTTG 675

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATCACTTCTCACAACGAAGACTACACTATTGTTGAACAATACGAAAGAACTGAAGGTAGA 660

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCTAAGAAGCCAGTTCAATTGCCAGGTTACTACTACGTTGACTCTAAGTTGGAC 600

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATCCACAAGGCCTTGAAGTTGAAGGACGGTGGTCACTACTTGGTCGAATTCAAGTCTATC 540

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGGTTGGGAAGCTTCTACCGAACGTTTGTACCCAAGAGACGGTGTCCTTGAAGGGTGAA 480

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTTAAGTTCATCGGTGTTAACTTCCCATCTGACGGTCCAGTTATGCAAAAGAAGACT 420

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGTGTTGTTACTGTTACTCAAGACTCTTCTTTGCAAGACGGTTGTTTCATTTAC 300

Qy   81 AspTyrLysLysLeuSerPhePheGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GACTACAAGAAGTTGTCTTTCCCAGAAGGTTTCAAGTGGGAAAGAGTTATGAACTTCGAA 240

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TTGTCTCCACAATTCCAATACGGTTCTAAGGTTTACGTTAAGCACCCAGCTGACATTCCA 180

Qy   41 HisAsnThrValThrLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CACAACACTGTTAAGTTGAAGGTTACTAAGGGTGGTCCATTGCCTTTTGCTTGGGACATT 120

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGAGAGTCTTCAAGAACGTTATTAAGGAGTTACTAAGGGTGGTCCATTG 60

Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
```

US-10-006-922A-12 (1-225) x ABA95922 (1-678)

XX
DE Yeast optimised RFP related DNA SEQ ID NO 18.
XX
KW Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW Escherichia coli; green fluorescent protein; biotechnology; gene; ss.
XX
OS Anthozoa.
XX
FH
FT Key   Location/Qualifiers
FT CDS   1..678
FT   /*tag= a
FT   /product= "yeast optimised Red Fluorescent Protein"
XX
XX
XX
PN DE20001395-U1.
XX
PD 15-MAR-2001.
XX
PF 27-JAN-2000; 2000DE-02001395.
XX
PR 27-JAN-2000; 2000DE-02001395.
XX
PA (GPCB-) GPC BIOTECH AG.
XX
DR WPI; 2002-228394/29.
DR P-PSDB; ABB08834.
XX
PT New DNA encoding red fluorescent protein, useful as marker in
PT biotechnology, has sequence optimized for expression in eukaryotes,
PT especially yeast or plants.
XX
PS Disclosure; Page 14; 19pp; German.
XX
CC The invention relates to DNA (I) containing either sequence ABA95905 or
CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC plants, especially dicotyledonous plants including Nicotiana tabacum or
CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC especially Escherichia coli. RFP is useful in the same way as green
CC fluorescent protein but is more generally applicable in modern
CC biotechnology. (I) are optimised for expression in yeast and so generate
CC fluorescent protein at higher levels with stronger fluorescence and thus
CC RFP at higher levels with stronger fluorescence and thus lowers the
CC detection limit and gives a better signal-to-noise ratio. The present
CC sequence is that of a polynucleotide encoding the yeast optimised RFP,
CC useful to the invention
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.3e-140
Score: 1214.00   Length: 678
Percent Similarity: 100.00%   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 6   Mismatches: 0
DB: Indels: 0
                     Gaps: 0

US-10-006-922A-12 (1-225) x ABA95921 (1-678)

```
Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCAAGGTTCGCATGGAAGGA  60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGGTCAATGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAGAGGC  120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTCGCTTGGGATATT  180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   181 TTGTCACCACAATTTCAGTATGGAAGCAAGAGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTATAAAAAGCTGTCATTTCCTGAAGGAATTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCAGTTATGCAAAAGAAGACA 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGGCTGGGAAGCCAGGACACTGAGCGTCTTTGTATCCCTGGTGATGGCGTTGAAAGGGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATTCATAAGGCTTCTGAAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGAAGCCTGTGCAGCTACCAGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATAACAAGCCACCAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGCCGC 660

Qy   221 HisHisHisLeuPheLeu 225
          ||||||||||||||||||
Db   661 CACCATCTGTTCCTT 675

RESULT 7
ABA95920
ID ABA95920 standard; RNA; 678 BP.
XX
AC ABA95920;
XX
DT 29-MAY-2002 (first entry)
XX
DE Yeast optimised RFP encoding RNA SEQ ID NO 16.
XX
KW Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW Escherichia coli; green fluorescent protein; biotechnology; ss.
XX
OS Anthozoa.
XX
PN DE20001395-U1.
XX
PD 15-MAR-2001.
XX
PF 27-JAN-2000; 2000DE-02001395.
XX
PR 27-JAN-2000; 2000DE-02001395.
XX
PA (GPCB-) GPC BIOTECH AG.
XX
DR WPI; 2002-228394/29.
XX
PT New DNA encoding red fluorescent protein, useful as marker in
PT biotechnology, has sequence optimized for expression in eukaryotes,
PT especially yeast or plants.
XX
PS Disclosure; Page 13; 19pp; German.
XX
CC The invention relates to DNA (I) containing either sequence ABA95905 or
CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC plants, especially dicotyledonous plants including Nicotiana tabacum or

Left column

```
CC  Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC  especially Escherichia coli. RFP is useful in the same way as green
CC  fluorescent protein but is more generally applicable in modern
CC  biotechnology. (II) are optimized for expression in yeast and so generate
CC  RFP at higher levels with stronger fluorescence and thus lowers the
CC  detection limit and gives a better signal-to-noise ratio. The present
CC  sequence is that of an RNA sequence corresponding to the yeast optimised
CC  RFP encoding cDNA given as SEQ ID NO 1 (ABA95905)
XX
SQ  Sequence 678 BP; 198 A; 147 C; 159 G; 0 T; 174 U; 0 Other;
```

US-10-006-922A-12 (1-225) x ABA95920 (1-678)

```
Alignment Scores:
Pred. No.:
Score:                2.3e-140
Percent Similarity:   1214.00
Best Local Similarity: 100.00%
Query Match:          100.00%
                      6
```

```
Length:        678
Matches:       225
Conservative:    0
Mismatches:      0
Indels:          0
Gaps:            0
```

```
Qy    1  MetArgSerIleLysAsnValIleLeuGlyGluPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  AUGAGAUCUAUCAAGAACGUCAUCUUAGGGGAAUUCAUGAGAUUCAAGGUUAGAAUGGAAGGU  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACUGUUAACGGUCACGAAUUCGAAAUUGAAGGUGAAGGUGAAGGUGAGACCAUACGAAGGU  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAACACUGUCAAGUUGAAGGUUACUAAGGGUGGUCCAUUGCCAUUCGCUUGGGACAUC  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  UUGUCUCCACAAUUCCAAUACGGUUCUAAGGUCUACGUCAAGCACCCAGCUGACAUUCCA  240

Qy   81  AspThrTyrLysLeuLysPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACUACAAGAAGUGUCCUUCCAGAAGGUUUCAAGUGGGAAAGAGUCAUGAACUUCGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGUGGUGUUGUUACUGUUACUCAAGACUCUUCCUUGCAAGACGGUUGUUUCAUCUAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGUCAAGUUCAUCGGUGUCAACUUCCCAUCUGACGGUCCAGUGAUGCAAAAGAAGACU  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  AUGGGUUGGGAAGCUUCUACCGAACGUUUGUACCCAAGAGACGGUGUCUUGAAGGGUGAA  480

Qy  161  IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  AUCCACCAAGGCCUUGAAGUUGAAGGACGGUGGUCACUACUUGGUCGAAUUCAAGUCUAUC  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  UACAUGGCUAAGAAGCCAGUUCAAUUGCCAGGUUACUACUACGUUGACUCUAAGUUGGAC  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  AUCACUUCUCACAACGAAGACUACACGAUCGUUGAACAGUACGAACGUACUGAAGGUAGA  660

Qy  221  HisHisHisLeuPheLeu  225
         ||||||||||||||||||
Db  661  CACCACUUGUUCUUG  675
```

Right column

```
AC  AAD46278;
XX
DT  27-DEC-2002 (first entry)
XX
DE  Discosoma sp. drFP583 (NFP-6) wild-type protein encoding DNA.
XX
KW  Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW  fluorescence activated cell sorting application; fluorescent timer;
KW  biosensor; fluorescence resonance energy transfer application; FRET;
KW  colouring agent; recombinant DNA application; analyte detection assay;
KW  sunscreen; second messenger detector; drFP583 protein; NFP-6; gene; ds.
XX
OS  Discosoma sp.
XX
FH  Key          Location/Qualifiers
FT  CDS          1..678
FT               /*tag= a
FT               /product= "drFP583 wild-type protein"
XX
PN  WO200268459-A2.
XX
PD  06-SEP-2002.
XX
PF  20-FEB-2002; 2002WO-US005749.
XX
PR  21-FEB-2001; 2001US-0270983P.
PR  04-DEC-2001; 2001US-00006922.
XX
PA  (CLON-) CLONTECH LAB INC.
XX
PI  Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
PI  Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
PT  New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT  of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT  analyte detection assays or fluorescence activated cell sorting
PT  applications.
XX
PS  Disclosure; Page 70; 80pp; English.
XX
CC  The invention relates to nucleic acid molecules encoding non-aggregating
CC  chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC  useful in analyte detection assays, as colouring agents, as markers in
CC  recombinant DNA applications, as sunscreens or filters, in fluorescence
CC  resonance energy transfer (FRET) applications, as biosensors in
CC  prokaryotic and eukaryotic cells, in screening assays, as second
CC  messenger detectors, in fluorescence activated cell sorting applications,
CC  in protease cleavage assays or as fluorescent timers. The present
CC  sequence is a DNA encoding Discosoma sp. drFP583 (NFP-6) wild-type
CC  protein of the invention
XX
SQ  Sequence 678 BP; 146 A; 223 C; 203 G; 106 T; 0 U; 0 Other;
```

RESULT 8
ID AAD46278 standard; DNA; 678 BP.
XX

US-10-006-922A-12 (1-225) x AAD46278 (1-678)

```
Alignment Scores:
Pred. No.:
Score:                2.3e-140
Percent Similarity:   1214.00
Best Local Similarity: 100.00%
Query Match:          100.00%
                      6
```

```
Length:        678
Matches:       225
Conservative:    0
Mismatches:      0
Indels:          0
Gaps:            0
```

```
Qy    1  MetArgSerIleLysAsnValIleLeuGlyGluPheMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGCGCTCTCCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGACGGC  120
```

Qy   41 HisAsnThrValLysGlyLeuLysGlyValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGGCGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGC 660

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db   661 CACCACCTGTTCCTG 675

DR   WPI; 2002-154595/20.
DR   P-PSDB; AAB17540.
XX
Example 1; Fig 1; 89pp; English.
XX
PS   The invention relates to a fluorescent timer protein having an emission
      spectrum that changes over time after synthesis from a first wavelength
      to a second wavelength. The fluorescent timer proteins are useful in
      monitoring the activity of a promoter, determining the age of a protein,
      identifying an agent that modulates the activity of a promoter and in
      enriching a population of cells comprising a fluorescent timer protein.
      The fluorescent timer proteins are also useful for assessing gene
      expression during development of a multicellular organism or during
      cellular differentiation, in response to a drug or other inducer of
      promoter activity, as a reporter to serve as a read-out of promoter
      activity, monitoring intracellular protein movement or translocation,
      protein trafficking, or protein stability, to investigate temporal
      aspects of the activity of a regulatory element, for determining cell
      fate during development and organ remodelling, in spatial and temporal
      visualisation of newly synthesized proteins and accumulated proteins, and
      in distinguishing between newly formed and pre-existing structures, e.g.
      membrane junctions and extracellular matrix components. The fluorescent
      timer proteins may further be used to investigate structures where photobleaching
      techniques are employed, as detectable labels, as selectable markers, as
      biosensors in prokaryotic and eukaryotic cells, in protease cleavage
      assays, and as second messenger detectors. The nucleic acids can be used
      to generate transgenic, non-human plants or animals or site-specific gene
      modifications in cell lines. The present sequence is a DNA encoding
      Discosoma sp. humanised wild-type Anthozoa protein drFP583 used for
      generating fluorescent proteins
XX
SQ   Sequence 678 BP; 146 A; 223 C; 203 G; 106 T; 0 U; 0 Other;

RESULT 9
AAD28207
ID   AAD28207 standard; DNA; 678 BP.
XX
AC   AAD28207;
XX
DT   22-APR-2002 (first entry)
XX
DE   Discosoma sp. humanised wild-type Anthozoa protein drFP583 DNA.
XX
KW   Fluorescent timer protein; protein movement; translocation; trafficking;
KW   promoter activity; gene expression; transgenic plant; gene modification;
KW   protein age; anthozoa protein; drFP583; ds.
XX
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /tag= a
FT   /product= "Humanised wild-type Anthozoa protein drFP583"
XX
PI   Pradkov AF, Terskikh A;

Alignment Scores:
Pred. No.:   2.3e-140
Score: 1214.00
Percent Similarity: 100.00%   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 6   Conservative: 0
DB: Mismatches: 0
                                       Indels: 0
                                       Gaps: 0

US-10-006-922A-12 (1-225) x AAD28207 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysGlyLeuLysGlyValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGGCGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

```
Db        361  AAGGTGAAAGTTCATCGGCGTGAAACTTCCCTCGACGCGCCCGTGATGCAGAAAGAACC
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlu 160
Qy        421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAG 480
          481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540
Db
Qy        161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db        541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600
Qy
Db        201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660
Qy
Db        221  HisHisHisLeuPheLeu 225
          661  CACCACCACCTGTTCCTG 675
```

```
RESULT 10
ID    ADC24127       standard; DNA; 678 BP.
AC    ADC24127;
XX
DT    18-DEC-2003 (first entry)
XX
DE    Discosoma wild-type red fluorescent protein DNA #1.
XX
KW    Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW    fluorescent protein variant; transcription induction detection;
KW    fluorescence energy resonance transfer; FRET; protein kinase;
KW    protein phosphatase; ion indicator; ds.
XX
OS    Discosoma.
XX
PN    US20030059835-A1.
XX
PD    27-MAR-2003.
XX
PF    10-APR-2002; 2002US-00121250.
PR    26-FEB-2001; 2001US-00794308.
PR    24-MAY-2001; 2001US-00866538.
XX
PT    Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT    variant having a reduced propensity to oligomerize, useful for detecting
PT    transcriptional activity.
XX
PA    (TSIE/) TSIEN R.Y.
PA    (CAMP/) CAMPBELL R.E.
XX
PI    Tsien RY, Campbell RE;
XX
DR    WPI; 2003-743764/70.
DR    P-PSDB; ADC24126.
XX
PS    Example 1; SEQ ID NO 2; 67pp; English.
XX
CC    The invention describes a polynucleotide sequence (I) encoding a
CC    Discosoma red fluorescent protein (DsRed) variant having a reduced
CC    propensity to oligomerise, comprising amino acid substitutions at the AB
CC    and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC    amino acids, given in the specification, where the substitutions result
CC    in reduced propensity of the DsRed variant to form tetramers. (I) is
CC    useful for detecting transcriptional activity by providing a host cells
CC    containing a vector which comprises (I) operatively linked to an
CC    expression control sequence, and an unit to assay fluorescence of the variant
CC    protein fluorescence, and assaying fluorescence of the variant
```

```
CC    fluorescent protein produced by (VII), where variant fluorescent protein
CC    fluorescence is indicative of transcriptional activity. A polynucleotide
CC    encoding a fusion protein is useful for the analysis of in vivo
CC    localisation or trafficking of a polypeptide of interest. A polypeptide
CC    marker is useful as markers to identify the location and amount of a
CC    target protein produced, where the target protein is fused to the marker,
CC    as a complement to or alternative for the green fluorescent protein or
CC    its spectral variant, for detecting induction of transcription, in
CC    applications involving fluorescence energy resonance transfer (FRET),
CC    which detects events as the function of the movement of fluorescent
CC    donors and acceptors towards or away from each other, for making
CC    fluorescent sensors for protein kinase and phosphatase activities or
CC    indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC    presence of a molecule in a sample, for identifying a specific
CC    interaction of a first and second molecule, for determining whether a
CC    sample contains an enzyme or for determining the pH of the sample. (I) is
CC    useful for identifying a region or condition that regulates the activity
CC    of an expression control sequence. This sequence encodes Discosoma wild-
CC    type red fluorescent protein.
XX
SQ    Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

```
US-10-006-922A-12 (1-225) x ADC24127 (1-678)

Alignment Scores:
Pred. No.:                                1
Score:                           2.3e-140
Score:                           1214.00          Length:             678
Percent Similarity:              100.00%          Matches:            225
Best Local Similarity:           100.00%          Conservative:         0
Query Match:                     100.00%          Mismatches:           0
DB:                                    10         Indels:               0
                                                  Gaps:                 0
```

```
Qy     1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCAAGGTTAGGATGGAAGGA 60

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ACGGTCAATGGACACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGTACCAGAAGGC 120

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CACAATACCGTAAAGCTAAAGGTTACCAAGGGGGACCTTTGCCATTTGGTTGGGATATT 180

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ATGGGCTGGGAGGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ATTCATAAGGCTCTGAAGCTTAAGGATGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TACATGGCAAAGAAGCCTGTCCAGCTACCACAGGGTACTACTTATGTGGACTCCAAACTGGAT 600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
```

Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAAGAACCGAGGACGC 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 221 HisHisLeuPheLeu 225

Db 661 CACCATCTGTTCCTT 675

Db RESULT 11
ID ADF70404
AC ADF70404;
    ADP70404 standard; cDNA; 678 BP.
XX
DT 12-FEB-2004 (first entry)
XX
DE Discosoma wild-type GFP variant cDNA SeqID27.
XX
KW ligand; orphan receptor protein; fusion protein; fluorescent protein;
KW cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
KW GFPuv; Enhanced GFP; EGFP; gene; 88.
XX
OS Discosoma sp.
XX
PN WO200307l272-Al.
XX
PD 28-AUG-2003.
XX
PF 21-FEB-2003; 2003WO-JP001901.
XX
PR 22-FEB-2002; 2002JP-0004572B.
PR 23-JUL-2002; 2002JP-002l3949.
PR 11-OCT-2002; 2002JP-00298237.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX
DR WPI; 2003-697654/66.
DR P-PSDB; ADF70403.
XX
PT Transformation of cells with a fusion protein of an orphan receptor
PT protein with a fluorescent protein useful for identification of ligands
PT to the orphan receptor.
XX
PS Disclosure; SEQ ID NO 27; 594pp; Japanese.
XX
CC This invention relates to a novel method of identifying ligands to an
CC orphan receptor protein which comprises transforming cells with DNA
CC encoding a fusion protein of the orphan receptor with a fluorescent
CC protein, so that the fusion protein is expressed in the cells (or cell
CC membranes isolated from them) and contacting the cells with the potential
CC ligand to be tested. A suitable fluorescent protein for incorporation in
CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,
CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
CC identification of ligands binding to an orphan receptor protein.
XX
SQ Sequence 678 BP; 204 A; 129 C; 179 G; 166 T; 0 U; 0 Other;

Db Discosoma sp.

Alignment Scores:
Pred. No.:   11
Score: 1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 10

US-10-006-922A-12 (1-225) x ADF70404 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40

Db 61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGAAGGCCATACGAAGGC 120
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 41 HisAsnThrValLeuLysValValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

Db 121 CACAATACCGTAAAAGTCGTTACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db 181 TTGTCACCACAATTCCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Db 301 GACGGTGGCGTCGTTACGGTCACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCAGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db 481 ATTCATAAGGCTCTGAAGCTCAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Db 541 TACATGGCCAAAAAGCCGGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220

Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 660

Qy 221 HisHisLeuPheLeu 225

Db 661 CACCATCTGTTCCTT 675

Db RESULT 12
ID ADL46204
AC ADL46204;
    ADL46204 standard; DNA; 678 BP.
XX
DT 20-MAY-2004 (first entry)
XX
DE Discosoma red fluorescent protein (DsRed) coding sequence.
XX
KW da; gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS Discosoma sp.
XX
FH Key   Location/Qualifiers
FT CDS   1..678
FT   /*tag= a
FT   /product= "DsRed protein"
XX
PN WO2003086446-Al.
XX
PD 23-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010879.
XX
PR 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-00209208.
XX
PA (REGC) UNIV CALIFORNIA.

```
XX
PI   Telen RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
DR   P-PSDB; ADL46203.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotide, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PN
PS
XX
CC
CC   Disclosure; SEQ ID NO 2; 166pp; English.
XX
CC
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially reduced tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   coding sequence.
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                1214.00              Length:            678
Percent Similarity:   100.00%              Matches:           225
Best Local Similarity: 100.00%             Conservative:      0
Query Match:          11                   Mismatches:        0
DB:                                        Indels:            0
                                           Gaps:              0

                   2.3e-140

US-10-006-922A-12 (1-225) x ADL46204 (1-678)

Qy       1 MetArgSerLysLysAsnValIleLeuHisGluPheMetArgPheLysValArgMetGluGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy      21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy      41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy      61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCGGCAGATATCCA 240

Qy      81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GACTATAAAAAGCTGTCATTCCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy     101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GACGGCGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy     121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAGGTCAAGTTCATTGGCGTGAACTTCCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy     141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy     161 IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATTCATGGAGGTCTCGAAGCTGAAAGACACGGTGGTGTCATACCTAGTTGAATTCAAAAGTATT 540

Qy     181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy     201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy     221 HisHisLeuPheLeu 225
           |||||||||||||||
Db     661 CACCATCTGTTCCTT 675
```

```
RESULT 13
ID   ADN33978 standard; DNA; 678 BP.
ADN33978
XX
AC   ADN33978;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Wild-type DsRED encoding sequence.
KW   Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED;
KW   ds.
OS   Discosoma sp.
XX
PN   WO200305158-A2.
XX
PD   03-JUL-2003.
XX
AC   18-DEC-2002; 2002WO-US040539.
PF
PR   19-DEC-2001; 2001US-0341723P.
PT   Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT   mutant of a Cnidarian chromo- or fluorescent protein or its mutant,
PT   useful for applications involving chromo- or fluorescent proteins.
PA   (UYCH-) UNIV CHICAGO.
XX
PS   Claim 8; SEQ ID NO 1; 65pp; English.
XX
CC
CC   The present invention relates to nucleic acid that encodes a rapidly
CC   maturing chromo or fluorescent mutant of a Cnidarian chromo- or
CC   fluorescent protein or its mutant. The protein is useful in applications
CC   involving nucleic acid encoding a chromo and/or fluorescent protein which involves
CC   useful for producing a chromo and/or fluorescent protein which involves
CC   growing the cell, whereby the protein is expressed, and isolating the
CC   protein substantially free of other proteins. The protein is useful in
CC   applications involving chromo- or fluorescent protein and is useful as
CC   PCR primers, hybridization probes, etc. The expression cassettes are
CC   useful for synthesizing related proteins. The chromoproteins are useful
CC   as coloring agents which are capable of imparting color or pigment to a
CC   particular composition of matter e.g. food compositions, pharmaceuticals,
CC   cosmetics, living organisms, e.g., animals and plants. The chromoproteins
CC   may also find use as labels in analyte detection assays, e.g. assays for
CC   biological analytes of interest and as selectable markers in recombinant
CC   DNA applications, e.g. the production of transgenic cells and organisms.
CC   The fluorescent proteins find use in a variety of different applications,
CC   e.g. in fluorescence resonance energy transfer (FRET) applications, as
CC   biosensors in prokaryotic and eukaryotic cells, in applications involving
CC   the automated screening of arrays of cells expressing fluorescent
CC   reporting groups by using microscopic imaging and electronic analysis, as
CC   second messenger detectors, and in fluorescence activated cell sorting,
CC   applications and as in vivo marker in animals. The fluorescent proteins
```

Bevis B, Glick B;

CC also find use in protease cleavage assays. The proteins can also be used
CC is assays to determine the phospholipid composition in biological
CC membranes and as a fluorescent timer. The present sequence represents the
CC wild-type DsRED encoding sequence.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADN33978 (1-678)

Alignment Scores:
Pred. No.:
Score: 2.3e-140
                     1214.00
Percent Similarity: 100.00%   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 11   Conservative: 0
DB: Mismatches: 0
                                   Indels: 0
                                   Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTTATGAGGTTTAAGGTTCGATGGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            |||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            ||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrThrSerHisAsnGluAspTyrThrIleValGluIleGluGlnTyrGluArgThrGluGlyArg 220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAACGCCACACAACGAAGACTATACAATCGTTGAGCGTGGAAGAACCGAGGGACGC 660

Qy 221 HisHisHisLeuPheLeu 225
            |||||||||||||
Db 661 CACCATCTGTTCCTT 675

DE Discosoma sp. red fluorescent protein (RED) DNA.
XX
FH Key   Location/Qualifiers
FT CDS   1..678
FT   /*tag= a
FT   /product= "Discosoma sp. red fluorescent protein (RED)"
XX
PN US2003203355-A1.
XX
PD 30-OCT-2003.
XX
PF 24-APR-2002; 2002US-00132067.
XX
PR 24-APR-2002; 2002US-00132067.
XX
PA (LALA-) LOS ALAMOS NAT LAB.
PA (REGC) UNIV CALIFORNIA.
XX
PI Bradbury AM, Zeytun A, Waldo GS;
XX
DR WPI; 2004-154325/15.
DR P-PSDB; ADI36421.
XX
PT Novel binding ligand with intrinsic fluorescence and comprising
PT fluorescent protein having heterologous binding sites, useful for
PT detecting target molecule.
XX
PS Example 6; SEQ ID NO 3; 23pp; English.
XX
CC The invention relates to binding ligands (fluorobodies) with intrinsic
CC fluorescence, which comprises green fluorescent protein (GFP) having
CC heterologous binding sites. The binding ligand is useful for detecting
CC the target molecule and is efficiently detects the target molecule. The
CC present sequence is Discosoma sp. red fluorescent protein (RED) DNA used
CC in the exemplification of the invention.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADI36420 (1-678)

Alignment Scores:
Pred. No.:
Score: 2.3e-140
                     1214.00
Percent Similarity: 100.00%   Length: 678
Best Local Similarity: 100.00%   Matches: 225
Query Match: 12   Conservative: 0
DB: Mismatches: 0
                                   Indels: 0
                                   Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTTATGAGGTTTAAGGTTCGATGGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            |||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

```
Qy  101  AspGlyGlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCCTGTGCACGTTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221  HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db  661  CACCATCTGTTCCTT 675
```

```
RESULT 15
ADM97768
ID   ADM97768 standard; DNA; 678 BP.
XX
AC   ADM97768;
XX
DT   01-JUL-2004  (first entry)
XX
DE   D sp red fluorescent protein coding sequence SEQ ID NO: 21.
XX
KW   dsj; gene; enzyme; sensor cell; fluorescent protein;
KW   signal transduction detection system; promoter; targeting sequence;
KW   targeted drug.
XX
OS   Discosoma sp.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..678
FT                  /*tag= a
FT                  /product= "fluorescent protein"
XX
SQ   Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

```
US-10-006-922A-12 (1-225) x ADM97768 (1-678)

Alignment Scores:
Pred. No.:               2.3e-140
Score:                   1214.00         Length:        678
Percent Similarity:      100.00%         Matches:       225
Best Local Similarity:   100.00%         Conservative:  0
Query Match:             100.00%         Mismatches:    0
DB:                      12              Indels:        0
                                         Gaps:          0
```

```
Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGGAAGGC 120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTCCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCCTGTGCACGTTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy  221  HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db  661  CACCATCTGTTCCTT 675
```

```
RESULT 16
AAA48743
ID   AAA48743 standard; cDNA; 695 BP.
XX
```

```
CC   comprises providing a homogeneous population of cells, where each of the
CC   cells comprises a signal transduction detection system and introducing
CC   into the population of cells an isolated DNA construct comprising a
CC   promoter operatively linked to a targeting sequence. The method is useful
CC   in developing a sensor cell for determining the activity of a target gene
CC   in the cell. The sensor cell and the methods are useful in developing new
CC   and therapeutic drugs directed to the targets. The present sequence is a
CC   coding sequence shown in the exemplification of the invention.
XX
SQ
```

```
DR   WO2004031415-A2.
XX
PN   WO2004031415-A2.
PD   15-APR-2004.
XX
PF   05-SEP-2003; 2003WO-US028078.
XX
PR   05-SEP-2002; 2002US-0408297P.
XX
PA   (VERT-) VERTEX PHARM INC.
PI   Whitney MA,  Zeh K,  Sanders PS;
XX
PT   Developing a sensor cell, useful in determining the activity of a target
PT   gene and in developing therapeutic drugs, comprises providing cells
PT   comprising a signal transduction detection system and introducing DNA
PT   construct into cells.
XX
PS   Disclosure; Page 167-168; 234pp; English.
XX
CC   The present invention relates to a method of developing a sensor cell,
CC   for determining the activity of a target gene in the cell, which
```

```
DR   WPI; 2004-330208/30.
DR   P-PSDB; ADM97769.
```

```
AC   AAA48743;
XX
DT   19-SEP-2000  (first entry)
XX
DE   Humanised Discosoma sp. "red" novel fluorescent protein drFP583 cDNA.
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling; ss.
XX
OS   Discosoma sp; "red".
OS   Synthetic.
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029473.
PR   11-DEC-1998; 98US-00210330.
PR   14-OCT-1999; 99US-00418529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 6; Page 75-76; 86pp; English.
XX
CC   The present sequence is humanised drFP583 cDNA. drFP583 is a full-length
CC   cDNA encoding a novel fluorescent protein (nFP) from Discosoma sp. "red",
CC   a non-bioluminescent species of the Class Anthozoa. The wild-type drFP583
CC   nucleotide sequence was altered to optimise the codons for expression of
CC   the fluorescent protein in mammalian cells. Fluorescent proteins can be
CC   used in fluorescent labeling, a useful tool for marking a protein, cell
CC   or organism of interest. Unlike other markers used in protein labeling,
CC   such as beta-galactosidase and luciferase, fluorescent proteins do not
CC   require an exogenous cofactor or substrate. Methods involving fluorescent
CC   proteins are also less laborious and less difficult to control than the
CC   traditional methods of fluorescent labeling, where a protein of interest
CC   is purified and then covalently conjugated to a fluorophore derivative.
CC   Novel fluorescent proteins isolated from species of the Class Anthozoa
CC   can be used as markers for gene expression and protein localization
CC   studies, and in fluorescence resonance energy transfer (FRET) reactions.
CC   They may have improved properties and better suitability for larger
CC   excitations compared to prior art fluorescent proteins such as green
CC   fluorescent protein
XX
SQ   Sequence 695 BP; 149 A; 228 C; 209 G; 109 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                  2.39e-140   1214.00    Length:          695
Percent Similarity:     100.00%                Matches:         225
Best Local Similarity:  100.00%                Conservative:    0
Query Match:            3                       Mismatches:      0
DB:                                             Indel:           0
                                                Gaps:            0

US-10-006-922A-12 (1-225) x AAA48743 (1-695)

Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 ATGCGGCTCTCCAAGAACGTCATCAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC    69

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   129

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle    60
```

```
Db  130 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   189

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro    80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  190 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCAGACATCCCC   249

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  250 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   309

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  310 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC   369

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  370 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC   429

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  430 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG   489

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  490 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC   549

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  550 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC   609

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleGluGlnTyrGluArgThrGluGlyArg   220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  610 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCGAGCAGTACGAGCGCACCGAGGGCCGC   669

Qy  221 HisHisLeuPheLeu 225
          |||||||||||||||
Db  670 CACCACCTGTTCCTG 684

RESULT 17
AAL47952
ID   AAL47952 standard; DNA; 859 BP.
XX
AC   AAL47952;
XX
DT   26-SEP-2002  (first entry)
XX
DE   Discosoma red fluorescent protein coding sequence.
XX
KW   Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
KW   modified yeast strain; environmental pollution; gene; ds.
XX
OS   Discosoma sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             54..731
FT                   /tag= a
FT                   /product= "red fluorescent protein"
XX
PN   DE1061872-A1.
XX
PD   20-JUN-2002.
XX
PF   12-DEC-2000; 2000DB-01061872.
PR   12-DEC-2000; 2000DB-01061872.
XX
PI   Lichtenberg-Frate H.
XX
PA   (LICH/) LICHTENBERG-FRATE H.
XX
DR   WPI; 2002-539633/58.
DR   P-PSDB; AAO18270.
```

XX

PT Modified yeast strain, useful for detecting toxic compounds in
PT environment, contains integrated cassettes responsive to genotoxic and
PT cytotoxic compounds.
XX
PS
XX
Disclosure; Page 20-21; 34pp; German.

The present invention relates to a modified yeast strain that contains,
integrated stably and functionally in its genome, a genotoxicity cassette
and a cytotoxicity cassette, each comprising a promoter and reporter
gene, both of which are different in the two cassettes. The modified
yeast strain is used to detect environmental pollution, especially
genotoxic and/or cytotoxic substances in complex environmental
contaminants, especially organic compounds, but also (non-)ionising
radiation and chemical carcinogens. Particular applications are in
monitoring (waste) water (e.g. as an early warning system), medical
toxicology screening and for industrial process control. The present
sequence is a marker gene suitable for use in the cassettes of the
present invention

SQ   Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;

US-10-006-922a-12 (1-225) x AAL47952 (1-859)

Alignment Scores:
Pred. No.:
Score: 3.22e-140
Percent Similarity: 1214.00   Length: 859
Percent Similarity: 100.00%   Matches: 225
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: Indels: 0
DB: 6   Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTTATGAGGTTTAAGGTTCGCATGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   54 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGGCGATACGAAGGC 173

Qy   41 HisGlnThrAlaLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   174 CACAATACCCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 293

Qy   81 AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   294 GACTAAAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy   121 LysValAlaLysProPheGlyAspAlaSerValGlyProValMetGlnLysLysThr 140
Db   414 AAGGTACAAGTTCATTGGGCGTGAAACTTTCCTTCCGATGGAACCTGTTTATGCAAAAGACA 473

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTGTTGAAAGGAGAG 533

Qy   161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   534 ATTCATCATGCGCTCAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   594 TACATCGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATAGTGACTCCAAACTGGAT 653

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   201 ... 220

Db   654 ATAACAAGGCACCACGAAGACTATACAAATCGTTGAGCAGTATGAAAGAAGACCGGAGGGACGC 713
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db   714 CACCACCTGTTCCTT 728

RESULT 18

ADY51715
ID   ADY51715 standard; DNA; 859 BP.
XX
AC   ADY51715;
XX
DT   05-MAY-2005 (first entry)
XX
DE   Discosoma sp. red fluorescent protein (RFP) DNA Seq 11.
XX
KW   fluorescence; mutagenesis; red fluorescent protein; gene; ds;
KW   protein interaction.
XX
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   54..731
FT   /*tag= a
FT   /product= "RFP protein"
XX
PN   WO200268605-A2.
XX
PD   06-SEP-2002.
XX
PF   26-FEB-2002; 2002WO-US006063.
XX
PR   26-FEB-2001; 2001WO-US006063.
PR   24-MAY-2001; 2001US-00794308.
PR   2001US-00865538.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Tsien RY, Baird GS, Campbell RE, Zacharias DA;
XX
DR   WPI; 2002-713372/77.
DR   P-PSDB; ADY51716.
XX
PT   New non-oligomerizing fluorescent protein containing at least one
PT   mutation that reduces or eliminates the ability of the protein to
PT   oligomerize, useful for making better and new assays for molecular
PT   biology.
XX
PS   Disclosure; SEQ ID NO 11; 117pp; English.

This invention relates to a novel non-oligomerizing fluorescent protein.
Specifically, it refers to the presence of at least one mutation in the
fluorescent protein that reduces or eliminates the ability of the protein
to oligomerize. The present invention describes fluorescent proteins
derived from naturally occurring green or red fluorescent proteins and
provides a fusion protein that comprises a non-oligomerizing fluorescent
protein linked to at least one protein of interest. As such, these fusion
proteins can be used in methods and compositions to determine the pH of a
sample, or whether the sample contains an enzyme, molecule or agent that
regulates the activity of an expression control sequence. Furthermore,
they may be used to identify a specific interaction of molecules, such
that they are useful for improving or developing new assays in the field
of molecular biology. This polynucleotide is the DNA sequence that
encodes the Discosoma sp. red fluorescent protein (RFP) of the invention.
XX
SQ   Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 3.22e-140
                    1214.00   Length: 859
Percent Similarity: 100.00%   Matches: 225
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: Indels: 0

```
DB:

US-10-006-922A-12 (1-225) x ADY51715 (1-859)

Qy                                                        7      Gaps:        0

Db          1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly 20
              ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTGCATGGAAGGA 113

Qy         54 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTGCATGGAAGGA 113

Qy         21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db        114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGAGGGAGGCCATACGAAGGC 173

Qy         41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db        174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy         61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db        234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy         81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db        294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy        101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db        354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy        121 LysValGlyPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db        414 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy        141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db        474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 533

Qy        161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db        534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy        181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db        594 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy        201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db        654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy        221 HisHisLeuPheLeu 225
Db        714 CACCATCTGTTCCTT 728
```

RESULT 19
AAD53432

ID   AAD53432 standard; DNA; 859 BP.
XX
AC   AAD53432;
XX
DT   28-MAY-2003   (first entry)
XX
DE   Discosoma species red fluorescent protein (RFP) encoding DNA.
XX
KW   Phosphorylation indicator; fluorescent protein; detection; phosphatase;
KW   kinase; red fluorescent protein; RFP; gene; ds.
XX
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   54..731
FT   /*tag= a
FT   /product= "Discosoma sp. red fluorescent protein (RFP)"
XX
PN   WO200295058-A2.

XX
PD   28-NOV-2002.
XX
PF   24-MAY-2002; 2002WO-US016955.
XX
PR   24-MAY-2001; 2001US-00865291.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Tsien RY, Ting AY, Zhang J;
XX
DR   WPI; 2003-148474/14.
DR   P-PSDB; AAB34962.
XX
PT   Novel chimeric phosphatase indicators, useful for detecting
PT   kinase/phosphatase in samples, has donor molecule, phosphorylatable
PT   domain, phosphoaminoacid binding domain, and acceptor molecule, in
PT   operative linkage.
XX
PS   Disclosure; Col 64-65; 38pp; English.
XX
CC   The present invention relates to chimeric phosphorylation indicators
CC   comprising a phosphorylation polypeptide and a fluorescent protein or in
CC   operative linkage, a donor molecule, a phosphorylatable domain, a
CC   phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The
CC   phosphorylation indicators of the invention are useful for detecting
CC   kinases or phosphatases in a biological sample. They are also useful in
CC   high throughput analysis e.g. for detecting a kinase inhibitor or
CC   phosphatase inhibitor. The present sequence is Discosoma species red
CC   fluorescent protein (RFP) encoding DNA used in the invention
XX
SQ   Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 3.22e-140
Percent Similarity: 1214.00   Length: 859
Best Local Similarity: 100.00%   Matches: 225
Query Match: 100.00%   Conservative: 0
DB: 8   Mismatches: 0
                                 Indels: 0
                                 Gaps: 0

```
US-10-006-922A-12 (1-225) x AAD53432 (1-859)

Qy          1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         54 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTGCATGGAAGGA 113

Qy         21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGAGGGAGGCCATACGAAGGC 173

Qy         41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
              ||||||||||||||||||||||||||||||||||||||||||||
Db        174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy         61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
              ||||||||||||||||||||||||||||||||||||||||
Db        234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy         81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
              ||||||||||||||||||||||||||||||||||||||||
Db        294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy        101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
              ||||||||||||||||||||||||||||||||||||||||
Db        354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy        121 LysValGlyPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
              ||||||||||||||||||||||||||||||||||||||||
Db        414 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy        141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
              ||||||||||||||||||||||||||||||||||||||||
```

```
Db   474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGGATNGGGGCGTTGTTGAAAGGAGAG 533
             ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Qy   161 IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 ATTCATATAGGCTCGAAGCTGAAAGACGGTGGTCATTACCTAGTGAATTCAAAAGTATT 593

Qy   181 TyrMetAlaLysGlyLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 TACATGGCAAAGGGAAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy   221 HisHisLeuPheLeu 225
             |||||||||||||||
Db   714 CACCATCTGTTCCTT 728
```

RESULT 20

Alignment Scores:

Pred. No.:	
Score:	3.22e-140
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	10

Length: 859
Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AAD61969 (1-859)

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   174 CACAAATACCGTAAAAGTTAAGGTAACCAAGGGGGACCTTTGCCCAATTGCTTGGGATATT 233

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   234 TTGTCACCACAANTTCAGTATGGAAGCAAGGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   294 GACTATAAAAAAGCTGTCATTCCTGAAGGAATTAAATGGGAAGGGTTCATGAACTTTGAA 353

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA 473

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGGATNGGGGCGTTGTTGAAAGGAGAG 533

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 ATTCATATAGGCTCGAAGCTGAAAGACGGTGGTCATTACCTAGTGAATTCAAAAGTATT 593

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   174 CACAAATACCGTAAAAGTTAAGGTAACCAAGGGGGACCTTTGCCCAATTGCTTGGGATATT 233

Qy   161 IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 ATTCATATAGGCTCGAAGCTGAAAGACGGTGGTCATTACCTAGTGAATTCAAAAGTATT 593

Qy   181 TyrMetAlaLysGlyLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 TACATGGCAAAGGGAAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy   221 HisHisLeuPheLeu 225
             |||||||||||||||
Db   714 CACCATCTGTTCCTT 728
```

RESULT 21

```
FT
FT                 /*tag= a
                   /product= "DsRed red fluorescent protein"
XX
PN   US20050026234-A1.
XX
PD   03-FEB-2005.
XX
PF   28-MAY-2004; 2004US-00857622.
PR   31-JAN-1996; 96US-00594575.
PR   31-JAN-1997; 97US-00792553.
PR   13-SEP-1999; 99US-00396003.
PR   24-MAY-2001; 2001US-00865291.
XX
XX
PA   (VIOL/) VIOLIN J D.
PA   (NEWT/) NEWTON A.C.
PA   (TSIE/) TSIEN R Y.
PA   (ZHAN/) ZHANG J.
XX
PT   Violin JD, Newton AC, Tsien RY, Zhang J;
XX
PS   Disclosure; SEQ ID NO 11; 68pp; English.
XX
DR   WPI; 2005-141377/15.
DR   P-PSDB; ADX26534.
XX
PT   New chimeric phosphorylation indicator comprises a first fluorescent
     protein, phosphoaminoacid binding domain with an FHA2 sequence and protein
     kinase C-phosphorylatable domain, useful for detecting protein kinase C
     or phosphates.
XX
CC   The invention relates to a chimeric phosphorylation indicator (CPI) (I)
CC   comprising in operative linkage, a first fluorescent protein, a
CC   phosphoaminoacid binding domain with an FHA2 sequence defined in the
CC   specification, a protein kinase C (PKC)-phosphorylatable domain, and a
CC   second fluorescent protein. The first and second fluorescent proteins are
CC   chosen from green fluorescent proteins (GFPs) and red fluorescent
CC   proteins (RFPs). The first and second fluorescent proteins exhibit a
CC   detectable resonance energy transfer, when the first fluorescent protein
CC   is excited. The PKC-phosphorylatable domain and phosphoaminoacid binding
CC   domain do not substantially emit light to excite the second fluorescent
CC   protein. (I) is useful for detecting a protein kinase C (PKC) or
CC   phosphates in a sample (such as biological sample e.g. cell, tissue
CC   sample, or their extracts). (I) is useful for detecting a kinase
CC   inhibitor or phosphatase inhibitor. The method of the invention is
CC   adapted to high throughput analysis. The current sequence represents the
CC   coding sequence of Discosoma DsRed RFP.
XX
SQ   Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;
```

```
US-10-006-922A-12 (1-225) x ADX26533 (1-859)

Alignment Scores:
Pred. No.:
Score:                3.22e-140
Score:                1214.00        Length:         859
Percent Similarity:   100.00%        Matches:        225
Best Local Similarity: 100.00%       Conservative:   0
Query Match:          14             Mismatches:     0
DB:                                  Indels:         0
                                     Gaps:           0

Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGC 173

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db  234 TTGTCACCACAATTTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCGACATACCA 293

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGACGGGTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCGGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGGGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleGlnGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAAGGGACGC 713

Qy  221 HisHisLeuPheLeu 225
          |||||||||||||||
Db  714 CACCATCTGTTCCTT 728
```

```
RESULT 22
ID   ACA62995                ACA62995 standard; DNA; 3311 BP.
XX
AC   ACA62995;
XX
DT   23-SEP-2003  (first entry)
XX
DE   Plasmid DNA containing coding sequence for RFP.
XX
KW   Fluorescently-tagged enzyme; substrate; cell population; GFP;
KW   quantification of enzymatic activity; green fluorescent protein;
KW   red fluorescent protein; RFP; enhanced green fluorescent protein; EGFP;
KW   enzymatic process; cellular process; cellular enzyme; chemotherapeutic drug;
KW   multidrug resistance; MDR; ds
XX
OS   Aequorea victoria.
OS   Synthetic.
XX
PN   US20030495974-A1.
XX
PD   13-MAR-2003.
XX
PI   Simon SM, Chen Y;
XX
DR   WPI; 2003-555145/52.
XX
PR   01-MAR-2001; 2001US-00797496.
PR   01-MAR-2001; 2001US-00797496.
XX
PA   (SIMO/) SIMON S M.
PA   (CHEN/) CHEN Y.
XX
PT   Simon SM, Chen Y;
XX
PT   Simultaneously quantifying in situ the relationship between an enzyme and
PT   its substrate for study of enzymatic processes at a cellular level;
PT   comprises optical measurements on cells expressing the fluorescently
PT   tagged enzyme.
```

XX
PS
XX
CC The present invention relates to a method for simultaneously quantifying
CC in situ the relationship between a fluorescently-tagged enzyme (E) and
CC its substrate in a population of cells. A population of cells expressing
CC (B) is created in which different cells contain different amounts of (E).
CC The cells are incubated with a substrate, and the concentration and
CC enzymatic activity of (E) in each cell are simultaneously quantified by
CC optical means. The fluorescently-tagged enzyme is produced by linking the
CC enzyme with Aequorea victoria green fluorescent protein (GFP), or its
CC variants or derivatives such as
CC enhanced green fluorescent protein (EGFP). The method of the invention is
CC useful for simultaneously quantifying in situ the relationship between an
CC enzyme and its substrate in a cell. The method is useful for the study of
CC enzymatic processes at the cellular level, and especially for examining
CC the interactions of cellular enzymes with chemotherapeutic drugs, e.g.
CC for elucidating mechanisms of multidrug resistance (MDR). The present
CC sequence represents a plasmid containing the coding sequence for RFP
XX
XX
SQ Sequence 3311 BP; 858 A; 778 C; 860 G; 815 T; 0 U; 0 Other;

Disclosure; Page 8-9; 21pp; English.

Alignment Scores:
Pred. No.:

Score: 2.15e-139	Length: 3311
Score: 1214.00	Matches: 225
Percent Similarity: 100.00%	Conservative: 0
Best Local Similarity: 100.00%	Mismatches: 0
Query Match: 100.00%	Indels: 0
DB: 9	Gaps: 0

US-10-006-922A-12 (1-225) x ACA62995 (1-3311)

Qy   1 MetArgSerSerLysSerAlaSerAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   289 ATGAGGTCTTCCAAGAAATGTTATTCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 348

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   349 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 408

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   409 CACAATACCGTAAAGTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 468

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   469 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 528

Qy   81 AspThrTyrLysLeuSerPheProGlyGlyPheLysTrpGluArgValMetAsnPheGlu 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   529 GACTATAAAAAGCTGTCATTTCCTGGAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 588

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   589 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 648

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   649 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 708

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   709 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 768

Qy   161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   769 ATTCATCATGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 828

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   829 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 888

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   889 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAAGGACGC 948

Qy   221 HisHisLeuPheLeu 225
           |||||||||||||||
Db   949 CACCACCTGTTCCTT 963

RESULT 23
ADC24134
ID ADC24134 standard; DNA; 681 BP.
XX
AC ADC24134;
XX
DT 18-DEC-2003 (first entry)
XX
DE Discosoma wild-type red fluorescent protein DNA #2.
XX
KW Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW fluorescent protein variant; transcription induction detection;
KW fluorescence energy resonance transfer; FRET; protein kinase;
KW protein phosphatase; ion indicator; ds.
XX
OS Discosoma.
XX
PN US200305935-A1.
XX
PD 27-MAR-2003.
XX
PI Tsien RY, Campbell RE;
XX
PF WPI; 2003-743764/70.
PF P-PSDB; ADC24126.
XX
PR 10-APR-2002; 2002US-00121258.
XX
PR 26-FEB-2001; 2001US-00794308.
PR 24-MAY-2001; 2001US-00865538.
XX
PA (TSIE/) TSIEN R Y.
PA (CAMP/) CAMPBELL R E.
XX
PT Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT variant having a reduced propensity to oligomerize, useful for detecting
PT transcriptional activity.
XX
PS Disclosure; SEQ ID NO 9; 67pp; English.
XX
CC The invention describes a polynucleotide sequence (I) encoding a
CC Discosoma red fluorescent protein (DsRed) variant having a reduced
CC propensity to oligomerise, comprising amino acid substitutions at the AB
CC and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC amino acids, given in the specification, where the substitutions result
CC in reduced propensity of the DsRed variant to form tetramers. (I) is
CC useful for detecting transcriptional activity by providing a host cells
CC containing a vector which comprises (I) operatively linked to an
CC expression control sequence, and an unit to assay the variant fluorescent
CC protein fluorescence, and assaying fluorescence of the variant
CC fluorescent protein produced by (VII), where variant fluorescent protein
CC fluorescence is indicative of transcriptional activity. A polynucleotide
CC encoding a fusion protein is useful for the analysis of in vivo
CC localisation or trafficking of a polypeptide of interest. A polypeptide
CC marker is useful as markers to identify the location and amount of a
CC target protein produced, where the target protein is fused to the marker,
CC as a complement to or alternative for the green fluorescent protein or
CC its spectral variant, for detecting induction of transcriptions, in
CC applications involving fluorescence energy resonance transfer (FRET),
CC which detects events as the function of the movement of fluorescent
CC donors and acceptors towards or away from each other, for making
CC fluorescent sensors for protein kinase and phosphatase activities or
CC indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC presence of a molecule in a sample, for identifying a specific
CC interaction of a first and second molecule, for determining whether a
CC sample contains an enzyme or for determining the pH of the sample. (I) is

CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This sequence represents a modified
CC   Discosoma wild-type red fluorescent protein coding region with humanised
CC   codon usage.
XX
SQ   Sequence 681 BP; 146 A; 223 C; 205 G; 107 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADC24134 (1-681)

Alignment Scores:
Pred. No.:
Score: 7.26e-140
Score: 1210.00   Length: 681
Percent Similarity: 100.00%   Matches: 224
Best Local Similarity: 99.56%   Conservative: 1
Query Match: 99.58%   Mismatches: 0
DB: 10   Indels: 0
                                       Gaps: 0

```
Qy   1 MetArgSerSerLysGluAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCCGCTTCTCCAAGAAGCTCATCAAGAAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy  41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy  81 AspThrTyrLysLeuSerPhePheGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GACACAGAAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||
Db 664 CACCACCTGTTCCTG 678
```

DE   Human codon optimised Discosoma red fluorescent protein (DsRed) DNA.
XX
KW   ds; gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
PN   WO2003086446-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010879.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-00209208.
XX
PA   (REGC) UNIV CALIFORNIA.
PI   Tsien RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PS   Disclosure; SEQ ID NO 3; 166pp; English.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers, and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   coding sequence with codons optimised for human expression.
XX
SQ   Sequence 681 BP; 146 A; 223 C; 205 G; 107 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADL46205 (1-681)

Alignment Scores:
Pred. No.:
Score: 7.26e-140
Score: 1210.00   Length: 681
Percent Similarity: 100.00%   Matches: 224
Best Local Similarity: 99.56%   Conservative: 1
Query Match: 99.67%   Mismatches: 0
DB: 11   Indels: 0
                                       Gaps: 0

```
Qy   1 MetArgSerSerLysGluAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCCGCTTCTCCAAGAAGCTCATCAAGAAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy  41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy  81 AspThrTyrLysLeuSerPhePheGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   coding sequence with codons optimised for human expression.
XX
SQ   Sequence 681 BP; 146 A; 223 C; 205 G; 107 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   7.26e-140   Length: 681
Score: 1210.00   Matches: 224
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 99.56%   Mismatches: 1
Query Match: 99.67%   Indels: 0
DB: 11   Gaps: 0

US-10-006-922A-12 (1-225) x ADL46225 (1-681)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluGluMetArgMetLysLeuTyrMetGluGly 20
        ::::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCGTTCCCCAGTTCCAGTTACGGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GACTACAACAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValGlyPheIleGlyValAsnPheProSerArgGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCAGACGGCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 664 CACCACCTGTTCCTG 678

RESULT 26
ID   AB224476 standard; DNA; 723 BP.
XX
AC   AB224476;

Db 244 GACTACAACAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303
CC
CC
CC
CC
CC
XX
SQ

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValGlyPheIleGlyValAsnPheProSerArgGlyProValMetGlnLysLysThr 140
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCAGACGGCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
Db 664 CACCACCTGTTCCTG 678

RESULT 25
ID   ADL46225 standard; DNA; 681 BP.
XX
AC   ADL46225;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Human codon optimised Discosoma red fluorescent protein (DsRed) DNA.
XX
KW   ds; gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
XX
PN   WO200306446-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010879.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-00209208.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Tsien RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PS   Disclosure; SEQ ID NO 23; 166pp; English.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable

XX
DT   25-MAR-2003 (first entry)
XX
DE   Mammalian codon optimised Discosoma red fluorescent protein DNA.
XX
KW   Mammalian codon optimised Discosoma red fluorescent protein; Discosoma;
KW   red fluorescent protein; directed evolution; gene; ds.
XX
OS   Discosoma sp.
OS   Mammalia.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..723
FT   /*tag= a
FT   /product= "Mammalian codon optimised Discosoma red
FT   fluorescent protein"
FT   /transl_except= (pos:616..618,aa:His)
XX
PN   WO20029499992-A2.
XX
PD   28-NOV-2002.
XX
PF   20-MAY-2002; 2002WO-US015968.
XX
PR   18-MAY-2001; 2001US-0291871P.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Peelle B;
XX
DR   WPI; 2003-120798/11.
DR   P-PSDB; ABP56678.
XX
PS   Example 1; Fig 1; 22pp; English.
XX
PT   New Discosoma red fluorescent protein, useful for functional screens as a
PT   reporter for gene transcription, for target characterization and
PT   localization of fusion proteins, or for scaffolds for protein and peptide
PT   libraries.
XX
CC   The present invention describes an isolated Discosoma red fluorescent
CC   protein (I) comprising a 241 residue amino acid sequence (see ABP56678,
CC   SI), with one or more point mutations at amino acid position N24, F125,
CC   K164, or M183. Also described: (1) a fusion protein comprising (I); (2)
CC   an isolated nucleic acid encoding (I); (3) a vector comprising the
CC   nucleic acid of (2); (4) a host cell comprising the vector; (5) a
CC   retroviral cDNA expression library comprising the nucleic acid of (2), or
CC   encoding (I); and (6) methods of making a fluorescent variant. The
CC   Discosoma red fluorescent proteins are useful for functional screens as a
CC   reporter for gene transcription (e.g. as a fusion protein), for target
CC   characterization and localisation of fusion proteins, or for scaffolds
CC   for protein and peptide libraries. The fluorescent proteins can also be
CC   used as selectable markers or reporter molecules for a variety of
CC   bioassays, including methods that use fluorescence activated cell sorting
CC   (FACS) as a selection mechanism. The method of directed protein evolution
CC   is useful for obtaining improved variants of red fluorescent protein. The
CC   variants of Discosoma red fluorescent protein have greatly improved
CC   brightness, expression, and/or folding kinetics as compared to wild type
CC   or a codon optimised variant. The present sequence encodes a mammalian
CC   codon optimised Discosoma red fluorescent protein, which is used in an
CC   example from the present invention.
XX
SQ   Sequence 723 BP; 161 A; 233 C; 221 G; 108 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 7.9e-140
                        1210.00
Percent Similarity: 100.00%
Percent Identity: 99.56%
Best Local Similarity: 99.67%
Query Match: 8

Length: 723
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x ABZ22476 (1-723)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
            :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCCCTTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCCATGGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCTCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGTATC 543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGCACCGAGGGCCGC 663

Qy   221 HisHisLeuPheLeu 225
            ||||||||||||||||
Db   664 CACCACCTGTTCCTG 678

RESULT 27
ADL18131
ID   ADL18131 standard; DNA; 1638 BP.
XX
AC   ADL18131;
XX
DT   06-MAY-2004 (first entry)
XX
DE   RFP:PS(NIa protease):AtOEP7:GFP fusion protein DNA SEQ ID NO:51.
XX
KW   chimeric protein; signal protein; trafficking signal targeting;
KW   proteolytic cleavage site; protease; protease inhibitor; gene; ds.
XX
OS   Arabidopsis thaliana.
OS   Tobacco vein mottling virus.
OS   Synthetic.
XX
PN   WO2003014381-A1.
XX
PD   20-FEB-2003.
XX
PF   08-AUG-2002; 2002WO-KR001515.
XX

PR   10-AUG-2001; 2001KR-00048123.
XX
PA   (AHRA-) AHRAM BIOSYSTEMS INC.
XX
PI   Hwang I, Kim DH, Lee YJ;
XX
DR   WPI; 2003-256596/25.
DR   P-PSDB; ADL18132.
XX
PT   New chimeric protein, useful for detecting protease inhibitors inside the
     cell or tissue.
XX
PS   Example 2; SEQ ID NO 51; 214pp; English.
XX
CC   The present invention describes a chimeric protein comprising at least
CC   one signal protein that has a trafficking signal targeting to a
CC   subcellular organelle and at least one proteolytic cleavage site for a
CC   protease. The chimeric protein is constructed, so that: (a) the
CC   trafficking signals of all the signal proteins are inactivated by linking
CC   the proteolytic site or a signal masking protein through the proteolytic
CC   site to the N-or C-terminus of the signal proteins, and so the chimeric
CC   protein is present in cytosol; (b) the trafficking signal of at least one
CC   signal protein is activated when the proteolytic cleavage site is cleaved
CC   by the protease, and as a result at least one fragment protein that
CC   includes the activated signal protein is transported to a subcellular
CC   organelle; and (c) the chimeric protein is labelled with at least one
CC   fluorescent protein and the position and intensity distribution of the
CC   fluorescent label signal in the cell is altered depending on the cleavage
CC   by the protease. Also described: (1) a chimeric protein which is constructed
CC   to express the chimeric protein in a cell; (2) a cell transformed with
CC   the recombinant gene or vector; (3) analysing the activity of a protease
CC   in vivo; (4) screening protease inhibitors in vivo; (5) a system for
CC   detecting a protease inside a cell; (6) a nucleic acid comprising the
CC   sequence encoding the chimeric protein for detecting protease activity in
CC   a cell; (7) a vector comprising the nucleic acid; (8) a kit for detecting
CC   a protease inside a cell comprising the chimeric protein or the vector;
CC   (9) detecting a protease inside a cell or tissue; and (10) detecting a
CC   protease inhibitor in vivo. The chimeric protein is useful for detecting
CC   protease inhibitors inside the cell or tissue. The present sequence
CC   encodes a fusion protein, which is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 1638 BP; 445 A; 450 C; 419 G; 324 T; 0 U; 0 Other;

     US-10-006-922A-12 (1-225) x ADL18131 (1-1638)

     Alignment Scores:
     Pred. No.:

     Score: 2.5e-139
     Score: 1210.00
     Percent Similarity: 100.00%
     Best Local Similarity: 99.56%
     Query Match: 99.67%
     DB: 10

                           Length: 1638
                           Matches: 224
                           Conservative: 1
                           Mismatches: 0
                           Indels: 0
                           Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGGGCGCTTCCTCCAAGAACGTCATCAAGGAGTTCATGCGGCTTCAAGGTGCGCATGGAGGGC 63

Qy   21 ThrValAlaAsnGlyHisGlyIlePheGluIleGluGlyGluGlyGlyAlaArgProTyrGluGly 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGGCGCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuGlyValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACACACCGTGAAGTTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACTACAAGAAGCTGTCTTCCCCGAGGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCGCGGACGGCGTGCTGAAGGGCGAG 483

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluGluTyrSerIle 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy   221 HisHisLeuPheLeu 225
            |||||||||||||||
Db   664 CACCACCTGTTCCTG 678

RESULT 28
ADL18155
ID   ADL18155 standard; DNA; 1647 BP.
XX
AC   ADL18155;
XX
DT   06-MAY-2004   (first entry)
XX
DE   RPP:PS(HIV-1 protease):AtOBP7:GFP fusion protein DNA SEQ ID NO:75.
XX
KW   chimeric protein; signal protein; trafficking signal targeting;
KW   proteolytic cleavage site; protease; protease inhibitor; gene; ds.
XX
OS   Arabidopsis thaliana.
OS   Human immunodeficiency virus 1.
OS   Synthetic.
XX
PN   WO2003014381-A1.
XX
PD   20-FEB-2003.
XX
DR   WPI; 2003-256596/25.
DR   P-PSDB; ADL18156.
XX
PR   10-AUG-2001; 2001KR-00048123.
PF   08-AUG-2002; 2002WO-KR001515.
XX
PA   (AHRA-) AHRAM BIOSYSTEMS INC.
XX
PI   Hwang I, Kim DH, Lee YJ;
XX
PT   New chimeric protein, useful for detecting protease inhibitors inside the
     cell or tissue.
XX
PS   Example 3; SEQ ID NO 75; 214pp; English.
XX
CC   The present invention describes a chimeric protein comprising at least
CC   one signal protein that has a trafficking signal targeting to a
CC   subcellular organelle and at least one proteolytic cleavage site for a
CC   protease. The chimeric protein is constructed, so that: (a) the
CC   trafficking signals of all the signal proteins are inactivated by linking
CC   the proteolytic site or a signal masking protein through the proteolytic

```
CC  site to the N- or C- terminus of the signal proteins, and so the chimeric
CC  protein is present in cytosol; (b) the trafficking signal of at least one
CC  signal protein is activated when the proteolytic cleavage site is cleaved
CC  by the protease, and as a result at least one fragment protein that
CC  includes the activated protease is a transported to a subcellular
CC  organelle; and (c) the chimeric protein is labelled with at least one
CC  fluorescent protein and the position and intensity distribution of the
CC  fluorescent label signal in the cell is altered depending on the cleavage
CC  by the protease. Also described: (1) a recombinant gene comprising a
CC  nucleic acid sequence encoding the chimeric protein which is constructed
CC  to express the chimeric protein in a cell; (2) a cell transformed with
CC  the recombinant gene or vector; (3) analysing the activity of a protease
CC  in vivo; (4) screening protease inhibitors in vivo; (5) a system for
CC  detecting a protease inside a cell; (6) a nucleic acid comprising the
CC  sequence encoding the chimeric protein for detecting protease activity in
CC  a cell; (7) a vector comprising the nucleic acid; (8) a kit for detecting
CC  (9) detecting a protease inside a cell or tissue; and (10) detecting a
CC  protease inhibitor in vivo. The chimeric protein is useful for detecting
CC  protease inhibitors inside the cell or tissue. The present sequence
CC  encodes a fusion protein, which is used in the exemplification of the
CC  present invention.
XX
SQ  Sequence 1647 BP; 450 A; 448 C; 421 G; 328 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADL18155 (1-1647)

Alignment Scores:
Pred. No.:
Score:                   2.52e-139
Percent Similarity:      1210.00
Best Local Similarity:   100.00%
Query Match:             99.56%
DB:                      99.67%
                         10

                         Length:              1647
                         Matches:             224
                         Conservative:        1
                         Mismatches:          0
                         Indels:              0
                         Gaps:                0

Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgMetGluGly 20
          ::|||||||||||||||||||||||||||||||||||||||||||||||
Db      4 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGTATGGAGGGC 63

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||
Db     64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAAGGCGAGGGC 123

Qy     41 HisAlaAsnThrValLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    124 CACAACACCGTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    184 CTGTCCCCCAGTTCCAGTACGGGTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 423

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValAlaGluGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    604 ATCACCTCCCACAACGAAGACTACACCATCGTGGCCGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy    221 HisHisHisLeuPheLeu 225
          ||||||||||||||||||
Db    664 CACCACCTGTTCCTG 678
```

```
Db
Qy
Db
Qy
```

```
RESULT 29
ADS75466
ID   ADS75466 standard; DNA; 2721 BP.
XX
AC   ADS75466;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Fibrohexamerin promoter, exon 1 and intron 1 synthetic sequence SEQ ID 1.
XX
KW   Spidroin; fibroin; silkworm; Ds-Red; Fibrohexamerin; gene; promoter; ds.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FH
FT   promoter        1..1452
FT                   /tag= a
FT                   /note= "Binding site for transcription factors of the
FT                   SGF1 forkhead type"
FT   TATA_signal     1420..1423
FT                   /tag= e
FT   misc_feature    1150..2026
FT                   /tag= b
FT                   /label= Fibrohexamerin promoter
FT   exon            1452..1525
FT                   /tag= f
FT                   /number= 1
FT                   /note= "Fibrohexamerin exon 1, encodes the signal
FT                   peptide"
FT   intron          1526..2026
FT                   /tag= g
FT                   /number= 1
FT                   /note= "Fibrohexamerin intron 1"
FT   CDS             2041..2721
FT                   /tag= h
FT                   /product= "Reporter protein Ds-Red"
FT   misc_signal     1379..1390
FT                   /tag= c
FT   CAAT_signal     1362..1365
FT                   /tag= d
XX
OS   Spidroin; fibroin; silkworm; Ds-Red; Fibrohexamerin; gene; promoter; ds.
```

PN   FR2852325-A1.
XX
PD   17-SEP-2004.
XX
PF   13-MAR-2003; 2003FR-00003137.
XX
PR   13-MAR-2003; 2003FR-00003137.
XX
PA   (CNRS) CNRS CENT NAT RECH SCI.
PA   (UYLY-) UNIV LYON 1 BERNARD CLAUDE.
PA   (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Chavancy G, Couble P, Durand B, Grenier A, Horard B, Julien E;
PI   Mauchamp B, Nony P, Prudhomme JC, Royer C;
XX
DR   WPI; 2004-670997/66.
XX
PT   New nucleic acid that directs expression of protein in silk glands of
PT   silkworm, useful for preparation of pharmaceutical proteins and for
PT   modifying textile properties of silk.
XX

PS   Claim 1; SEQ ID NO 1; 50pp; French.

XX

CC   The present invention relates to a novel nucleic acid (I) that directs
CC   expression of a protein of interest (II) specifically in the cells of the
CC   posterior silk-producing glands (A) of Bombyx mori. (I) comprises, in the
CC   5' to 3' direction: a regulatory region (RR) comprising signals for
CC   regulating expression of a polynucleotide of interest (III) in the
CC   specified cells and, under control of RR, a region that encodes a
CC   modified signal peptide (mSP). (I) is at least 90% identical with the
CC   1150-2026 nucleotide (nt) region of ADS75466, in which the codon
CC   containing nt 1486-1488 encodes Ala, Ile or Leu. (I) is optionally
CC   modified by insertion of 1-4 copies of the sequence ADS75467 (nucleotides
CC   1379-1390 of ADS75466), which is the binding site for transcription
CC   factors of the SGF1 forkhead type, functioning as activator of
CC   transcription and regulator of tissue-specific expression. (III) encodes
CC   a polypeptide, specifically spidroin (of spiders) or fibroin (from
CC   organisms of the genus Galleria). (II), also related expression cassettes
CC   and vectors, are used to prepare transgenic Bombyx that secrete (II)
CC   incorporated into silk threads, e.g. where (II) is of biomedical interest
CC   (e.g. hormones, antigens, enzymes, growth factors and receptors) or where
CC   the threads are intended for use as textiles and (II) modifies resistance
CC   to ductility or elasticity.

XX

SQ   Sequence 2721 BP; 748 A; 631 C; 606 G; 736 T; 0 U; 0 Other;

US-10-006-922a-12 (1-225) x ADS75466 (1-2721)

Alignment Scores:
Pred. No.:
Score: 5.11e-139
Score: 1210.00
Percent Similarity: 100.00%
Best Local Similarity: 99.56%
Query Match: 99.67%
DB: 13

Length: 2721
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

```
Qy      1 MetArgSerSerLysSerAlaAsnValIleLeuLysGluPheMetArgMetLysValGluGly  20
            :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2044 GTGGCGCTCTTCCAAGAAACGTCATCAAGGAGAGTTCATCGCGCTTCAAGTGCCGATGACGGC 2103

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2104 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2163

Qy     41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2164 CACACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2223

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2224 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2283

Qy     81 AspThrLysLysGlnLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2284 GACACCAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2343

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2344 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 2403

Qy    121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2404 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 2463

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2464 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCAGGGACGGCGTGCTGAAGGGCGAG 2523

Qy    161 IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2524 ATCCACCAcAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 2583

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2584 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2643

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluLleuLysGlyThrGlyPheAsnLysArg 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2644 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGACGTGCACCGAGGGCCGGC 2703

Qy    221 HisHisLeuPheLeu 225
            |||||||||||||||
Db   2704 CACCACCTGTTCCTG 2718
```

RESULT 30

ADS75468
ID   ADS75468 standard; DNA; 2772 BP.
XX
AC   ADS75468;
XX
DT   02-DEC-2004 (first entry)
XX
DE   Fibrohexamerin promoter, exon 1 and intron 1 synthetic sequence SEQ ID 3.
XX
KW   Spidroin; fibroin; silkworm; Ds-Red; Fibrohexamerin; gene; promoter; ds.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   promoter   1..1503
FT   /*tag= a
FT   /note= "Three copies of the binding site for
FT   transcription factors of the SGF1 forkhead type"
FT   TATA_signal   1471..1474
FT   /*tag= d
FT   exon   1503..1576
FT   /*tag= e
FT   CAAT_signal   1362..1365
FT   /*tag= b
FT   misc_signal   1371..1440
FT   /*tag= c
FT   intron   1577..2077
FT   /number= 1
FT   /*tag= f
FT   /note= "Fibrohexamerin intron 1"
FT   CDS   2092..2772
FT   /*tag= g
FT   /product= "Reporter protein Ds-Red"
FT   /note= "Fibrohexamerin exon 1, encodes the signal
FT   peptide"
FT   /number= 1
XX
PN   FR2853325-A1.
XX
PD   17-SEP-2004.
XX
PF   13-MAR-2003; 2003FR-00003137.
XX
PR   13-MAR-2003; 2003FR-00003137.
XX
PA   (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA   (UYLY-) UNIV LYON 1 BERNARD CLAUDE.
PA   (CNRS) CNRS CENT NAT RECH SCI.
XX
PI   Chavancy G, Couble P, Durand B, Grenier A, Horard B, Julien E;
PI   Mauchamp B, Nony P, Prudhomme JC, Royer C;
XX
DR   WPI; 2004-670997/66.
XX
PT   New nucleic acid that directs expression of protein in silk glands of
PT   silkworm, useful for preparation of pharmaceutical proteins and for
PT   modifying textile properties of silk.
XX
PS   Example 3; SEQ ID NO 3; 50pp; French.
XX
CC   The present invention relates to a novel nucleic acid (I) that directs

CC   expression of a protein of interest (II) specifically in the cells of the
CC   posterior silk-producing glands (A) of Bombyx mori. (I) comprises, in the
CC   5' to 3' direction: a regulatory region (RR) comprising signals for
CC   regulating expression of a polynucleotide of interest (III) in the
CC   specified cells and, under control of RR, a region that encodes a
CC   modified signal peptide (mSP). (I) is at least 90% identical with the
CC   1150-2026 nucleotide (nt) region of ADS75466, in which the codon
CC   containing nt 1486-1488 encodes Ala, Ile or Leu. (I) is optionally
CC   modified by insertion of 1-4 copies of the sequence ADS75467 (nucleotides
CC   1379-1390 of ADS75466), which is the binding site for transcription
CC   factors of the SGF1 forkhead type, functioning as activator of
CC   transcription and regulator of tissue-specific expression. (III) encodes
CC   a polypeptide, specifically spidroin (of spiders) or fibroin (from
CC   organisms of the genus Galleria), also related expression cassettes
CC   and vectors, are used to prepare transgenic Bombyx that secrete (II)
CC   incorporated into silk threads, e.g. where (II) is of biomedical interest
CC   (e.g. hormones, antigens, enzymes, growth factors and receptors) or where
CC   the threads are intended for use as textiles and (II) modifies resistance
CC   to ductility or elasticity. The present sequence is a modified version of
CC   ADS75466 which comprises three copies of the SGF1 forkhead type
CC   transcription factor binding site.
XX
SQ   Sequence 2772 BP; 760 A; 640 C; 609 G; 763 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADS75468 (1-2772)

Alignment Scores:
Pred. No.:
Score: 5.24e-139   Length: 2772
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
                    13   Indels: 0
                                        Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2095 GTGCGGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 2154

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2155 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2214

Qy   41 HisAsnThrValLysThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2215 CACAACACCGTGAAGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2274

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2275 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2334

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2335 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2394

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2395 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 2454

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2455 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 2514

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2515 ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2574

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2575 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 2634

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2635 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2694

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2695 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2754

Qy   221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   2755 CACCACCTGTTCCTG 2769

XX

RESULT 31
PN   AAL47954
PN   AAL47954 standard; DNA; 4692 BP.
XX
ID   DE10061872-A1.
XX
DR   WPI; 2002-539633/58.
XX
PT   Modified yeast strain, useful for detecting toxic compounds in
PT   environment, contains integrated cassettes responsive to genotoxic and
PT   cytotoxic compounds.
XX
PA   (LICH/) LICHTENBERG-FRATE H.
XX
PI   Lichtenberg-Frate H;
XX
PN   AAL47954;
XX
AC   AAL47954 standard; DNA; 4692 BP.
XX
DT   26-SEP-2002 (first entry)
XX
FH
PF   12-DEC-2000; 2000DE-01061872.
PR   12-DEC-2000; 2000DE-01061872.
XX
PS   Example 1; Page 24-25; 34pp; German.
XX
CC   The present invention relates to a modified yeast strain that contains,
CC   integrated stably and functionally in its genome, a genotoxicity cassette
CC   and a cytotoxicity cassette, each comprising a promoter and reporter
CC   gene, both of which are different in the two cassettes. The modified
CC   yeast strain is used to detect environmental pollution, especially
CC   genotoxic and/or cytotoxic substances in complex environmental
CC   contaminants, especially organic compounds, but also (non-)ionising
CC   radiation and chemical carcinogens. Particular applications are in
CC   monitoring (waste) water (e.g. as an early warning system), medical
CC   toxicology screening and for industrial process control. The present
CC   sequence is a vector suitable for use in the present invention
XX
OS   Synthetic.
XX
KW   Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
KW   modified yeast strain; environmental pollution; vector; ds.
DE   Modified yeast strain related vector pDsRed1-N1.
XX
SQ   Sequence 4692 BP; 1111 A; 1268 C; 1247 G; 1066 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AAL47954 (1-4692)

Alignment Scores:
Pred. No.:
Score: 1.1e-138   Length: 4692
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
                    6   Indels: 0
                                        Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   682 GTGCGGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 741

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40

```
Db  742 ACCGGTGAACGGCCACGGAGTTCGGAGATTCGAGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCCGCCCTACGAGGGC 801
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   41 HisAsnThrValLysGlyValThrLysGlyGlyGlyProLeuProPheAlaTrpAsnIle 60
Db  802 CACAACACCGTGAACGGCCGTGAAGGGTGACCAAGGGCGGCGGGCCCTTGCCCTGGGACATC 861

Qy   61 LeuSerProGlnPheGlnThrTyrGlySerLysValThrValValTyrValHisProAlaAspIlePro 80
Db  862 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGGAGCACCCCGCCGACATCCCC 921

Qy   81 AspTyrTyrLysGlyLeuSerPheProGluGlyTrpGluArgValMetAsnPheGlu 100
Db  922 GACTACTACAAGGGCGTCCTTCCCCGAGGGGTTCAAGTGGGAGAGCGGCGTTGATGAACTTCGAG 981

Qy  101 AspGlyGlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIleTyr 120
Db  982 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCAGGACGGCTGCTTCATCTAC 1041

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 1042 AAGGTGAAGTTCATCGGCGTGAACTTCCCGTAATGCAGAAGAAGACC 1101

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 1102 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGACGGCGTGCTGAAGGGCGAG 1161

Qy  161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 1162 ATCCACCACGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1221

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 1222 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1281

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleGlnGlnTyrGluArgThrGluGlyArg 220
Db 1282 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCCAGCAGTACGAGCGCACCGAGGGCCGC 1341

Qy  221 HisHisLeuPheLeu 225
Db 1342 CACCACCTGTTCCTG 1356
```

Alignment Scores:

Pred. No.:	1.1e-138	Length: 4692
Score:	1210.00	Matches: 224
Percent Similarity:	100.00%	Conservative: 1
Best Local Similarity:	99.56%	Mismatches: 0
Query Match:	99.67%	Indels: 0
DB:	10	Gaps: 0

US-10-006-922A-12 (1-225) x ACC44640 (1-4692)

```
Qy    1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682 GTGCGGCTTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATCGGAGGC 741

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 801

Qy   41 HisAsnThrValLysGlyValThrLysGlyGlyGlyProLeuProPheAlaTrpAsnIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802 CACAACACCGTGAAGGGTGACCAAGGGCGGCGGGCCCTTGCCCTGGGACATC 861

Qy   61 LeuSerProGlnPheGlnThrTyrGlySerLysValThrValValTyrValHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGGAGCACCCCGCCGACATCCCC 921

Qy   81 AspTyrTyrLysGlyLeuSerPheProGluGlyTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922 GACTACTACAAGGGCCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 981

Qy  101 AspGlyGlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCCTGCAGGACGGCTGCTTCATCTAC 1041

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAAATGCAGAAGAAGACC 1101

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGGCGTGCTGAAGGGCGAG 1161
```

Qy   161 IleHisGlyAlaLeuLysLysAlaSerGlyGlyHisHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1162 ATCCACAAGGCCCTGAAGCTGAAGAAGGCGGGCCACCACTACCTGGTGGAGTTCAAGTCCATC 1221

Qy   181 TyrMetAlaLysLysProValGlnIleuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1222 TACATGGCCAAGAAGCCCGTGCAGATCCCGGGTACTACTACTACGTGGACTCCAAGCTGGAC 1281

Qy   201 IleThrSerHisAsnGluAspGlyTyrThrIleValGlnGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1282 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 1341

Qy   221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   1342 CACCACCTGTTCCTG 1356

RESULT 33
AAD10003
ID   AAD10003 standard; DNA; 5436 BP.
XX
AC   AAD10003;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Plasmid pSPM6 to express goosecoid cDNA fused to red fluorescent protein.
XX
KW   Mouse; fertility; reproduction; gametogenesis; microinjection; CMV;
KW   cytomegalovirus; SV40; simian virus 40; infection; goosecoid gene;
KW   plasmid pSFM 6; embryogenesis; ds.
XX
OS   Mus musculus.
XX
OS   Cytomegalovirus.
OS   Rhesus macaque polyoma virus.
OS   Unidentified.
OS   Chimeric.
XX
PN   WO20014B224-A1.
XX
PD   05-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-AU001596.
XX
PR   24-DEC-1999; 99AU-00004884.
XX
PA   (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Thresher R, Hinds L, Hardy C, Whyard S, Vignarajan S, Grewe PM;
PI   Patil J;
XX
DR   WPI; 2001-425672/45.
XX
PT   Novel construct for preventing embryogenesis in animals comprises native
PT   promoter, blocking DNA which abrogates function of crucial gene and
PT   genetic switch to regulate expression/repression of blocker/gene
PT   knockout.
XX
PS   Example 13; Page 215-217; 241pp; English.
XX
CC   The invention relates to a construct which allows animals to be bred in
CC   captivity but renders them infertile in the wild by allowing reversible
CC   control over fertility and reproduction. The construct comprises a native
CC   promoter, a blocking DNA sequence contoured for and designed to abrogate
CC   a crucial gene's function or to cause its mis-expression, and a genetic
CC   switch to regulate controlled expression/repression of the blocker/gene
CC   knockout. The construct is useful for preventing embryogenesis or
CC   gametogenesis in animals by stably transforming an animal cell with the
CC   construct by microinjection, transfection or infection, where the
CC   construct stably integrates into the genome by homologous recombination,
CC   and implanting the cell into a host organism, where a whole animal
CC   develops from the implanted cell. The present sequence is plasmid pSFM 6
CC   containing mouse goosecoid promoter expressing goosecoid cDNA fused to
CC   red fluorescent protein. The plasmid contains CMV promoter and SV40 polyA
CC   and SV40 ori
XX
SQ   Sequence 5436 BP; 1247 A; 1525 C; 1492 G; 1172 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 1.35e-138
Percent Similarity: 1210.00
Best Local Similarity: 100.00%
Query Match: 99.56%
DB: 99.67%

Length: 5436
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x AAD10003 (1-5436)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
          :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1426 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCGGC 1485

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1486 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1545

Qy   41 HisAsnThrValLysLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1546 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1605

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1606 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1665

Qy   81 AspTyrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1666 GACTACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1725

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPhelleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1726 GACGGCGGCGTGGTGACCGTGACCCAAGACTCCTCCCTGCAGGACGGCTCGTTCATCTAC 1785

Qy   121 LysValLysProPheGlyAspAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1786 AAGGTGAAGTTCATCGGCGTGACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 1845

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1846 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGGGTGCTGAAGGGCGAG 1905

Qy   161 IleHisGlyAlaLeuLysLysAlaSerGlyGlyHisHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1906 ATCCACAAGGCCCTGAAGAAGGCGGGCCACCACTACCTGGTGGAGTTCAAGTCCATC 1965

Qy   181 TyrMetAlaLysLysProValGlnIleuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1966 TACATGGCCAAGAAGCCCGTGCAGATCCCGGGTACTACTACTACGTGGACTCCAAGCTGGAC 2025

Qy   201 IleThrSerHisAsnGluAspGlyTyrThrIleValGlnGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2026 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2085

Qy   221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   2086 CACCACCTGTTCCTG 2100

RESULT 34
ADE24111
ID   ADE24111 standard; DNA; 6893 BP.
XX
AC   ADE24111;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Proviral plasmid DNA SFr.
XX
KW   site-specific DNA recombination; Cre recombinase;

```
KW   protein-transduction domain; gene activation; gene inactivation;
KW   chromosomal translocation; reporter protein; ds; circular.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FH
FT   misc_RNA        1..435
FT                   /tag= a
FT                   /note= "plasmid pUC fragment"
FT   repeat_region   436..996
FT                   /tag= b
FT                   /standard_name= "LTR"
FT                   /note= "MPSV 5'-LTR (Delta-31)"
FT   primer_bind     997..1014
FT                   /tag= f
FT                   /note= "Unidentified primer"
FT   5'UTR           1015..1560
FT                   /tag= d
FT   misc_RNA        1561..1599
FT                   /tag= e
FT                   /note= "loxP1 sequence"
FT   CDS             1600..2289
FT                   /tag= c
FT                   /product= "DsRed1"
FT                   /note= "from Discosoma sp"
FT   misc_RNA        2290..2343
FT                   /tag= g
FT                   /note= "loxP2 sequence"
FT   CDS             2347..3150
FT                   /tag= h
FT                   /product= "eGFP"
FT                   /note= "from Aequorea victoria"
FT   CDS             3164..4019
FT                   /tag= i
FT   3'UTR           4020..4080
FT                   /tag= j
FT                   /product= "wPRE"
FT   repeat_region   4081..4641
FT                   /tag= k
FT                   /standard_name= "LTR"
FT                   /note= "3'-LTR"
FT   misc_RNA        4642..6893
FT                   /tag= l
FT                   /note= "plasmid pUC fragment"
XX
SQ   Sequence 6893 BP; 1564 A; 1940 C; 1778 G; 1611 T; 0 U; 0 Other;
```

```
CC   in eukaryotic cells, in vivo or in vitro, e.g. for partial inactivation
CC   or activation of genes, also to create a chromosomal translocation. The
CC   modified recombinase can enter intact (not electrically or chemically
CC   treated) cells when added to culture medium and catalyzes recombination
CC   events in the nucleus with recombination rate over 50%, after a single
CC   application. The Cre-recombinase does not need to be supplied from
CC   nucleic acid introduced into the cell, so its effect is immediate but
CC   short-lived, limiting cytotoxic effects; eliminating the risk of non-
CC   specific integration of DNA into the genome and providing quick results.
CC   This sequence represents the proviral plasmid Cre reporter construct Sfr
CC   used in the method described in the invention.
XX
```

US-10-006-922A-12 (1-225) x ADE24111 (1-6893)

```
Alignment Scores:
Pred. No.:                        1.89e-138
Score:                            1210.00
Percent Similarity:               100.00%        Length:              6893
Best Local Similarity:            99.56%         Matches:              224
Query Match:                      99.67%         Conservative:           1
DB:                               10             Mismatches:             0
                                                 Indels:                 0
                                                 Gaps:                   0
```

```
Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          :::|::::||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db   1603 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1662

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1663 ACCGTGAACGGCCACGAGTTCGAGATCGAAGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1722

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1782

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1842

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPhePheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1843 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1902

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1903 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1962

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATCGCAGAAGAAGACC 2022

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2023 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2082

Qy    161 IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2083 ATCCACGGCGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 2142

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2143 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 2202

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2203 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2262

Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   2263 CACCACCTGTTCCTG 2277
```

RESULT 35

```
CC   This invention describes a novel method of site-specific
CC   recombination in eukaryotic cells in vitro comprising using a
CC   bacteriophage P1 Cre recombinase that lacks heterologous protein-
CC   transduction domains. The invention also contains a reporter system for
CC   detecting site-specific DNA recombination in eukaryotic cells.
CC   Recombination with Cre recombinase lacking heterologous protein-
CC   transduction domains is used for therapeutic site-specific recombination
PS   Claim 23; SEQ ID NO 8; 84pp; German.
XX
PT   In vitro or in vivo site-specific DNA recombination, useful e.g. for gene
PT   inactivation, using Cre recombinase that lacks heterologous protein-
PT   transduction domain.
PN   WO200070931-A2.
PD   28-AUG-2003.
PF   19-FEB-2003; 2003WO-EP001680.
PR   21-FEB-2002; 2002DE-01007313.
PR   16-JUL-2002; 2002DE-01032196.
PA   (VISI-) VISION 7 GMBH.
PI   Baum C, Will E, Ostertag W, Klump H, Schiedlmeier B;
DR   WPI; 2003-767353/72.
```

AAD09979

ID   AAD09979 standard; DNA; 7910 BP.
XX
AC   AAD09979;
XX
DT   12-SEP-2001 (first entry)
XX
DE   pBiT(dHSP)-RFP-oHoxDS/BH construct for transfection of pacific oysters.
XX
KW   Simian virus 40; SV40; fertility; reproduction; gametogenesis; fruitfly;
KW   microinjection; infection; plasmid pBiT(dHSP)-RFP-oHoxDS/BH; oyster; HSP;
KW   embryogenesis; RFP; heat shock promoter; red fluorescent protein; ds.
XX
OS   Crassostrea gigas.
OS   Rhesus macaque polyoma virus.
OS   Drosophila melanogaster.
OS   Unidentified.
OS   Chimeric.
XX
PN   WO200148224-A1.
XX
PD   05-JUL-2001.
XX
PI   Thresher R, Hinds L, Hardy C, Whyard S, Vignarajan S, Grewe PM;
PI   Patil J;
XX
PF   22-DEC-2000; 2000WO-AU001596.
XX
PR   24-DEC-1999; 99AU-00004884.
XX
PA   (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PT   Novel construct for preventing embryogenesis in animals comprises native
PT   promoter, blocking DNA which abrogates function of crucial gene and
PT   genetic switch to regulate expression/repression of blocker/gene
PT   knockout.
XX
PS   Example 8; Page 193-196; 241pp; English.
XX
CC   The invention relates to a construct which allows animals to be bred in
CC   captivity but renders them infertile in the wild by allowing reversible
CC   control over fertility and reproduction. The construct comprises a native
CC   promoter, a blocking DNA sequence contoured for and designed to abrogate
CC   a crucial gene's function or to cause its mis-expression, and a genetic
CC   switch to regulate controlled expression/repression of the blocker/gene
CC   knockout. The construct is useful for preventing embryogenesis or
CC   gametogenesis in animals by stably transforming an animal cell with the
CC   construct by microinjection, transfection or infection, where the
CC   construct stably integrates into the genome by homologous recombination,
CC   and implanting the cell into a host organism, where a whole animal
CC   develops from the implanted cell. The present DNA sequence is plasmid
CC   construct pBiT(dHSP)-RFP-oHoxDS/BH used for transfection of Pacific
CC   oysters. The plasmid comprises oyster Hox gene under the under the
CC   control of Drosophila heat shock promoter (dHSP), red fluorescent protein
CC   and SV40 PolyA
XX
SQ   Sequence 7910 BP; 2016 A; 1893 C; 1924 G; 2077 T; 0 U; 0 Other;

Qy   1 MetArgSerGlyArgLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyGly 20
              ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2705 GTGCGGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 2646

RESULT 36
ABS56664

ID   ABS56664 standard; DNA; 9320 BP.
XX
AC   ABS56664;
XX
DT   24-JAN-2003 (first entry)
XX
DE   Plasmid p5-Puro-CMV-(N'-EGFP)-CMV-Red(EGFP-EJ) DNA.
XX
KW   Vector; genotoxicity; food component; cosmetic; medicine; cancer;
KW   genetic stability; hygromycin; ds.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   misc_feature   1..1592
FT   /*tag= a
FT   /note= "retroviral vector p5NM"
FT   CDS   1617..2216
FT   /*tag= b
FT   /product= "puromycin resistance gene from pRetroON"
FT   promoter   2267..2848
FT   /*tag= c
FT   /note= "CMV promoter from pEGFP-N1"
FT   CDS   2906..3348
FT   /*tag= d
FT   /product= "N'-EGFP"

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2645 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2586

Qy   41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2585 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2526

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2525 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2466

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2465 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2406

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2405 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 2346

Qy   121 LysValLysLeuArgGlyThrAsnPheProSerAspGlyProValMetGlnLysLysThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2345 AAGGTGAAGTTCCGGGTGACGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 2286

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2285 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2226

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2225 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 2166

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2165 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2106

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2105 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2046

Qy   221 HisHisLeuPheLeu 225
              |||||||||||||||
Db   2045 CACCACCTGTTCCTG 2031

US-10-006-922a-12 (1-225) x AAD09979 (1-7910)

Alignment Scores:
Pred. No.:
Score: 2.3e-138
Length: 7910
Matches: 224
Best Local Similarity: 99.56%
Percent Similarity: 100.00%
Conservative: 1
Percent Identity: 99.67%
Query Match: 4

Length: 7910
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

FT   misc_feature   3374..3392
FT   /*tag= e
FT   /note= "retroviral vector p5NM"
FT   promoter   3411..3992
FT   /*tag= f
FT   /note= "CMV promoter from pDsRed1-N1"
FT   CDS   4038..4718
FT   /*tag= g
FT   /product= "pDsRed1-N1"
FT   CDS   4766..5508
FT   /*tag= h
FT   /product= "EGFP-EJ"
FT   misc_feature   5527..9320
FT   /*tag= i
FT   /note= "retroviral vector p5NM"

XX
PD   12-SEP-2002.
XX
PN   WO200270740-A2.
XX
PR   05-MAR-2001; 2001DB-01010449.
XX
PR   28-FEB-2002; 2002WO-EP002194.
XX
PI   Wiesmueller L;
PA   (WIES/) WIESMUELLER L.
XX
PT   New vector containing recombination sequences, useful e.g. for assessing
PT   compounds for genotoxicity, cancer susceptibility and mutational
PT   frequency.

DR   WPI; 2002-682909/73.
XX
PS   Disclosure; Page 104-106; 106pp; German.
XX
CC   This invention describes a novel vector containing, at least, two
CC   homologous DNA sequences that, through DNA exchange, induce a change in a
CC   gene present in the vector. The products of the invention are used to
CC   assess genotoxicity of compounds (e.g. components of foods, cosmetics and
CC   medicine), to determine susceptibility for development or progression of
CC   cancer, for analysis of (non-)conservative recombination processes, end
CC   joining and inactivating/reverting mutations, for characterization of
CC   selected genes with respect to maintenance or reduction of genetic
CC   stability, and to determine genetic (in)stability and recombinational or
CC   mutational frequency of cells, tissues or organisms. The vector allows an
CC   analysis to be performed in hours (making it suitable for large-scale,
CC   routine use), and can be used with a variety of eukaryotic cells and
CC   animals. All known types of DNA recombinations can be examined. This
CC   sequence represents a plasmid p5-Puro-CMV-(N'-EGFP)-CMV-Red(EGFP-EJ) DNA,
CC   described in the disclosure of the invention

SQ   Sequence 9320 BP; 2124 A; 2687 C; 2474 G; 2035 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1
Score: 1210.00
Percent Similarity: 100.00%
Best Local Similarity: 99.56%
Query Match: 99.67%
DB:

US-10-006-922A-12 (1-225) x ABS56664 (1-9320)

Pred. No.:   1   Length: 9320
Score: 2.89e-138   Matches: 224
                                     Conservative: 1
                                     Mismatches: 0
                                     Indels: 0
                                     Gaps: 0

Qy   1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4041 GTGCGGCTCCTCAAGAACGTCATCAAGGAGTTCATGCGGCTTCAAGGTGTGCCATGGAGGGC 4100

Qy   21 ThrValAsnGlyHisGlyPheGlyIleGluGlyGluGlyAsnProTyrGluGlyGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4101 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCCCTACGAGGGC 4160

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4161 CACAGACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 4220

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4221 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 4280

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4281 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 4340

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4341 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 4400

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4401 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 4460

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4461 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 4520

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4521 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 4580

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4581 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 4640

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4641 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 4700

Qy   221 HisHisLeuPheLeu 225
           ||||||||||||||
Db   4701 CACCACCTGTTCCTG 4715

RESULT 37
AD104104
ID   AD104104 standard; DNA; 9658 BP.
XX
AC   AD104104;
XX
DT   22-APR-2004 (first entry)
XX
DE   Vector pTnMod (Oval/Red) - Quail.
XX
KW   ss; vector; transposase; promoter; insertion sequence; Kozak; Tn10;
KW   ovalbumin; vitellogenin; polyA; conalbumin; enhancer; ovomucoid;
KW   signal sequence; transgenic animal; Huntington's disease;
KW   alpha-1-antitrypsin deficiency; Alzheimer's disease; breast cancer;
KW   cystic fibrosis; galactosaemia; congenital hypothyroidism;
KW   maple syrup urine disease; neurofibromatosis; phenylketonuria;
KW   sickle cell disease; Smith-Lemli-Opitz Syndrome; autoimmune disease;
KW   shipping fever; cattle; mastitis; bacterial; viral; infection; Types I;
KW   type; II; diabetes.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   misc_feature   1..130
FT   /*tag= a
FT   /note= "Remainder of P1(-) on from pBluescript11 sk(-)"
FT   misc_feature   131..132
FT   /*tag= b
FT   /note= "Residue from ligation of restriction enzyme sites
FT   used in construction"
FT   promoter   133..1777
FT   /*tag= c
FT   /note= "CMV promoter/enhancer from pGWiz"
FT   misc_feature   1778..1779

```
FT
FT   CDS             1780..2987
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /*tag= d
FT                   /product= "transposase"
FT                   /note= "Modified from Tn10 by optimising codons for
FT                   stability"
FT                   /*tag= e
FT   misc_feature    2988..2993
FT                   /note= "Two engineered stop codons"
FT                   /*tag= f
FT   polyA_signal    2995..3410
FT                   /note= "Taken from pGWiz"
FT                   /*tag= g
FT   misc_feature    3415..3718
FT                   /note= "Non-coding DNA residual from pNK2859"
FT                   /*tag= h
FT   misc_feature    3719..3761
FT                   /note= "Non-coding lambda DNA residual from pNK2859"
FT                   /*tag= i
FT   insertion_seq   3762..3831
FT                   /note= "Left insertion sequence recognised by Tn10"
FT                   /*tag= j
FT   misc_feature    3832..3837
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /*tag= k
FT   misc_feature    3838..4044
FT                   /note= "Multiple cloning site from pBluescript11 sk(-)"
FT                   /*tag= l
FT   misc_feature    4045..4049
FT                   /note= "Includes steroid-dependant response element"
FT                   /*tag= m
FT   misc_feature    4050..4934
FT                   /note= "Japanese quail ovalbumin promoter"
FT                   /*tag= n
FT   promoter        4935..4942
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /*tag= o
FT   misc_signal     4943..5092
FT                   /note= "Japanese quail ovalbumin signal sequence"
FT                   /*tag= p
FT   misc_feature    5093..5098
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /*tag= q
FT   CDS             5099..5991
FT                   /note= "DsRed reporter coding sequence including polyA
FT                   from pDsRed1.1"
FT                   /*tag= r
FT   misc_feature    5992..5997
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /*tag= s
FT   misc_signal     5998..6036
FT                   /note= "Multiple cloning site from ZeroBlunt Topo cloning
FT                   vector"
FT                   /*tag= t
FT   misc_signal     6037..6042
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /*tag= u
FT   misc_signal     6043..6475
FT                   /note= "Multiple cloning site from pBluescript11 sk(-)"
FT                   /*tag= v
FT   misc_signal     6476..6480
FT                   /*tag= w
FT   misc_signal     6481..6550
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT   insertion_seq   6551..6592
FT                   /*tag= x
FT                   /note= "Right insertion sequence recognised by Tn10"
FT   misc_feature    6593..7457
FT                   /*tag= y
FT                   /note= "Non-coding lambda DNA residual from pNK2859"
FT   misc_feature    7458..9658
FT                   /*tag= z
FT                   /note= "Non-coding DNA residual from pNK2859"
FT   misc_feature
FT                   /*tag= aa
FT                   /note= "Remainder from pBluescript11 sk(-)"
XX
PN   WO2004003157-A2.
XX
PD   08-JAN-2004.
XX
PF   26-JUN-2003; 2003WO-US020389.
XX
PR   26-JUN-2002; 2002US-0392415P.
PR   21-JAN-2003; 2003US-0441377P.
PR   21-JAN-2003; 2003US-0441381P.
PR   21-JAN-2003; 2003US-0441392P.
PR   21-JAN-2003; 2003US-0414405P.
PR   21-JAN-2003; 2003US-0441447P.
PR   21-JAN-2003; 2003US-0441502P.
XX
PA   (TRAN-) TRANSGENRES LLC.
PA   (LOUI-) UNIV LOUISIANA STATE & AGRIC & MECH COLL.
XX
PI   Cooper RK, Cadd GG, Fioretti WC, Deboer KF;
XX
DR   WPI; 2004-083042/08.
XX
PT

PS   Example 4; SEQ ID NO 4; 150pp; English.

New vector comprising a transposase gene operably linked to a promoter,
useful in treating Alzheimer's disease, breast cancer, phenylketonuria,
autoimmune diseases, bacterial or viral infections or Types I and II
diabetes.

CC   This sequence represents a new vector, pTnMod, comprising a transposase
CC   gene operably linked to a first promoter and one or more gene of interest
CC   operably linked to one or more additional promoters, where the one or
CC   more genes of interest and their operably linked promoters are flanked by
CC   transposase insertion sequences recognised by the transposase and the
CC   first promoter comprises a modified Kozak sequence comprising ACCATG. The
CC   transposase is a Tn10 transposase and is modified in one to twenty of the
CC   first codons. The gene of interest in operably linked to a second
CC   promoter, which is a constitutive promoter or an inducible promoter, e.g.
CC   an ovalbumin or a vitellogenin promoter. The vector further comprises a
CC   polyA sequence, a conalbumin polyA sequence, operably linked to the
CC   transposase gene. The vector further comprises two stop codons operably
CC   linked to the transposase gene. The first or second gene of interest is
CC   operably linked to a second or third promoter, respectively. The first
CC   and a second gene of interest are operably linked to a second promoter.
CC   The vector further comprises an enhancer operably linked to the one or
CC   more genes of interest. The enhancer comprises at least a portion of an
CC   ovalbumin enhancer. The vector further comprises an egg directing
CC   targeting sequence, operably linked to the one or more genes of interest.
CC   The vector is useful in producing transgenic animals with desired
CC   proteins or molecules. They are also useful in treating Huntington's
CC   disease, alpha-1-antitrypsin deficiency, Alzheimer's disease, breast
CC   cancer, cystic fibrosis, galactosaemia, congenital hypothyroidism, maple
CC   syrup urine disease, neurofibromatosis, phenylketonuria, sickle cell
CC   disease, Smith-Lemli-Opitz Syndrome, autoimmune diseases, shipping fever
CC   in cattle, mastitis, bacterial or viral infections or Types I and II
CC   diabetes.
XX
SQ   Sequence 9658 BP; 2572 A; 2397 C; 2148 G; 2541 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.:            3.04e-138            Length:            9658
```

```
Score:                 1210.00      Matches:          224
Percent Similarity:    100.00%      Conservative:     1
Best Local Similarity:  99.56%      Mismatches:       0
Query Match:            99.67%      Indels:           0
DB:                    12           Gaps:             0

US-10-006-922A-12 (1-225) x ADI04104 (1-9658)

Qy    1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
          :::|||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db 5099 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 5158

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5159 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 5218

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5219 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5278

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5279 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5338

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5339 GACACATACAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 5398

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5399 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 5458

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5459 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 5518

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5519 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTCCTGAAGGGCGAG 5578

Qy  161 IleHisIleLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5579 ATCCACATCAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 5638

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5639 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 5698

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5699 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 5758

Qy  221 HisHisHisLeuPheLeu 225
          ||||||||||||||||||
Db 5759 CACCACCACCTGTTCCTG 5773
```

```
RESULT 38
ID   ADI04103    standard; DNA; 9678 BP.
XX
AC   ADI04103;
XX
DT   22-APR-2004  (first entry)
XX
DE   Vector pTnMod (Oval/Red)-Chicken.
XX
KW   88; vector; transposase; promoter; insertion sequence; Kozak; Tn10;
KW   ovalbumin; vitellogenin; polyA; conalbumin; enhancer; ovomucoid;
KW   signal sequence; transgenic animal; conalbumin;
KW   alpha-1-antitrypsin deficiency; Alzheimer's disease; breast cancer;
KW   cystic fibrosis; galactosaemia; congenital hypothyroidism;
KW   maple syrup urine disease; neurofibromatosis; phenylketonuria;
KW   sickle cell disease; Smith-Lemli-Opitz Syndrome; autoimmune disease;
KW   shipping fever; cattle; mastitis; bacterial; viral; infection; Types I;
KW   type; II; diabetes.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FH
FT   misc_feature    1..130
FT                   /note= "Remainder of F1(-) on from pBluescriptll sk(-)"
FT                   /tag= a
FT   misc_feature    131..132
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /tag= b
FT   promoter        133..1777
FT                   /note= "CMV promoter/enhancer from pGWiz"
FT                   /tag= c
FT   misc_feature    1778..1779
FT                   /note= "CMV promoter is modified by the addition of an
FT                   ACC sequence upstream of ATG"
FT                   /tag= d
FT   CDS             1780..2987
FT                   /product= "Transposase"
FT                   /note= "Modified from Tn10 by optimising codons for
FT                   stability"
FT                   /tag= f
FT   misc_feature    2988..2993
FT                   /note= "Two engineered stop codons"
FT                   /tag= g
FT   misc_feature    2994
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /tag= h
FT   polyA_signal    2995..3410
FT                   /note= "Taken from pGWiz"
FT                   /tag= i
FT   misc_feature    3415..3718
FT                   /note= "Non-coding DNA residual from pNK2859"
FT                   /tag= j
FT   misc_feature    3719..3761
FT                   /note= "Non-coding lambda DNA residual from pNK2859"
FT                   /tag= k
FT   insertion_seq   3762..3831
FT                   /note= "Left insertion sequence recognised by Tn10"
FT                   /tag= l
FT   misc_feature    3832..3837
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /tag= m
FT   misc_feature    3838..4044
FT                   /note= "Multiple cloning site from pBluescriptll sk(-)"
FT                   /note= "May be used to insert the sequence of interest
FT                   into the vector"
FT                   /tag= n
FT   misc_signal     4045..4049
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /tag= o
FT   misc_signal     4050..4951
FT                   /note= "Upstream elements"
FT                   /tag= p
FT   misc_signal     4960..5112
FT                   /note= "Chicken ovalbumin signal sequence"
FT                   /tag= q
FT   misc_feature    5113..5118
FT                   /note= "Residue from ligation of restriction enzyme sites
FT                   used in construction"
FT                   /tag= r
FT   CDS             5119..6011
FT                   /note= "DsRed reporter coding sequence including polyA
FT                   from pDsRed1.1"
```

```
FT   misc_feature   6012..6017
FT                  /*tag= s
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   misc_signal    6018..6056
FT                  /*tag= t
FT                  /note= "Multiple cloning site from ZeroBlunt Topo cloning
FT                  vector"
FT   misc_signal    6057..6062
FT                  /*tag= u
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   misc_signal    6063..6495
FT                  /*tag= v
FT                  /note= "Multiple cloning site from pBluescriptll sk(-)"
FT   misc_signal    6496..6500
FT                  /*tag= w
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   misc_feature   6501..6570
FT                  /*tag= x
FT                  /note= "Non-coding DNA residual from pNK2859"
FT   insertion_seq  6571..6612
FT                  /*tag= y
FT                  /note= "Right insertion sequence recognised by Tn10"
FT   misc_feature   6613..7477
FT                  /*tag= z
FT                  /note= "Non-coding lambda DNA residual from pNK2859"
FT   misc_feature   7478..9678
FT                  /*tag= aa
FT                  /note= "Remainder from pBluescriptll sk(-)"
XX
PN   WO2004003157-A2.
PD   08-JAN-2004.
XX
PF   26-JUN-2003; 2003WO-US020389.
XX
PR   26-JUN-2002; 2002US-03924415P.
PR   21-JAN-2003; 2003US-0441377P.
PR   21-JAN-2003; 2003US-0441381P.
PR   21-JAN-2003; 2003US-0441392P.
PR   21-JAN-2003; 2003US-0441405P.
PR   21-JAN-2003; 2003US-0441447P.
PR   21-JAN-2003; 2003US-0441502P.
XX
PA   (TRAN-) TRANSGENRES LLC.
PA   (LOUI) UNIV LOUISIANA STATE & AGRIC & MECH COLL.
XX
PI   Cooper RK, Cadd GG, Fioretti WC, Deboer KF;
XX
DR   WPI; 2004-083042/08.
XX
PT   New vector comprising a transposase gene operably linked to a promoter,
PT   useful in treating Alzheimer's disease, breast cancer, phenylketonuria,
PT   autoimmune diseases, bacterial or viral infections or Types I and II
PT   diabetes.
PS   Example 3; SEQ ID NO 3; 150pp; English.
XX
CC   This sequence represents a new vector, pTnMod (Oval/Red)-Chicken,
CC   comprising a transposase gene operably linked to a first promoter and one
CC   or more gene of interest operably linked to one or more additional
CC   promoters, where the one or more genes of interest and their operably
CC   linked promoters are flanked by transposase insertion sequences
CC   recognized by the transposase and the first promoter comprises a modified
CC   Kozak sequence comprising ACCATG. The transposase is a Tn10 transposase
CC   and is modified in one to twenty of the first codons. The gene of
CC   interest in operably linked to a second promoter, which is a constitutive
CC   promoter or an inducible promoter, e.g. an ovalbumin or a vitellogenin
CC   promoter. The vector further comprises a polyA sequence, a conalbumin
CC   promoter. The vector further comprises a polyA sequence, a conalbumin
CC   promoter, operably linked to the transposase gene. The vector
CC   further comprises two stop codons operably linked to the transposase
```

```
CC   gene. The first or second gene of interest is operably linked to a second
CC   or third promoter, respectively. The first and a second gene of interest
CC   are operably linked to a second promoter. The vector further comprises an
CC   enhancer operably linked to the one or more genes of interest. The
CC   enhancer comprises at least a portion of an ovalbumin or
CC   ovomucoid signal sequence or vitellogenin targeting sequence, operably
CC   linked to the one or more genes of interest. The vector is useful in
CC   producing transgenic animals with desired proteins or molecules. They are
CC   also useful in treating Huntington's disease, breast cancer, cystic fibrosis
CC   deficiency, Alzheimer's disease, alpha-1-antitrypsin
CC   galactosaemia, congenital hypothyroidism, maple syrup urine disease,
CC   neurofibromatosis, phenylketonuria, sickle cell disease, Smith-Lemli-
CC   Optiz Syndrome, autoimmune diseases, shipping fever in cattle, mastitis,
CC   bacterial or viral infections or Types I and II diabetes.
XX
SQ   Sequence 9678 BP; 2586 A; 2396 C; 2152 G; 2544 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AD104103 (1-9678)

Alignment Scores:
Pred. No.:
Score:                 1210.00          Length:        9678
Percent Similarity:    100.00%          Matches:       224
Best Local Similarity: 99.56%           Conservative:  0
Query Match:           12               Mismatches:    1
                                        Indels:        0
                                        Gaps:          0
                                        DB:

Qy     1 MetArgSerSerLysAlaAsnValIleLysGluGluMetArgPheLysValArgProTyrGluGly 20
           :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5179 ACCGTGAAACGGCGCCACGAAGTTCGAGATCGAGAGGGGGGCGGCGCCCCCTACGAGGGC 5238

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5119 GTGCGCGTCCTCCAAGAACGTCATCAAGGAGAGTTCATGGCGCTTCAAGGTGCCCATGGAGGGC 5178

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5239 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5298

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5299 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5358

Qy    81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5359 GACTACAAGAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 5418

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5419 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 5478

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5479 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 5538

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5539 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 5598

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5599 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 5658

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5659 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 5718

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluIleTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5719 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCTACGAGCGCACCGAGGGCCGC 5778

Qy   221 HisHisLeuPheLeu 225
```

```
Db   5779 CACCACCTGTTCCTG 5793

RESULT 39

ID   AD104102 standard; DNA; 10263 BP.
XX
AC   AD104102;
XX
DT   22-APR-2004 (first entry)
XX
DE   Vector pTnMod (CMV/Red).
XX
OS   Synthetic.
XX
KW   ss; vector; transposase; promoter; insertion sequence; Kozak; Tn10;
KW   ovalbumin; vitellogenin; polyA; conalbumin; enhancer; ovomucoid;
KW   signal sequence; transgenic animal; Huntington's disease;
KW   alpha-1-antitrypsin deficiency; Alzheimer's disease; breast cancer;
KW   cystic fibrosis; galactosaemia; congenital hypothyroidism;
KW   maple syrup urine disease; neurofibromatosis; phenylketonuria;
KW   sickle cell disease; Smith-Lemli-Opitz Syndrome; autoimmune disease;
KW   shipping fever; cattle; mastitis; bacterial; viral; infection; Types I;
KW   type; II; diabetes.
XX
FH   Key            Location/Qualifiers
FH
FT   misc_feature   1..130
FT                  /*tag= a
FT                  /note= "Remainder of F1(-) on from pBluescript11 sk(-)"
FT   misc_feature   131..132
FT                  /*tag= b
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   promoter       133..1777
FT                  /*tag= c
FT                  /note= "CMV promoter/enhancer from pGWiz"
FT   promoter       ;
FT                  /*tag= c
FT                  /note= "CMV promoter is modified by the addition of an
FT                  ACC sequence upstream of ATG"
FT   misc_feature   1778..1779
FT                  /*tag= d
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   CDS            1780..2987
FT                  /*tag= e
FT                  /product= "transposase"
FT                  /note= "Modified from Tn10 by optimising codons for
FT                  stability"
FT   misc_feature   2988..2993
FT                  /*tag= f
FT                  /note= "Two engineered stop codons"
FT   misc_feature   2994
FT                  /*tag= g
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   polyA_signal   2995..3410
FT                  /*tag= h
FT                  /note= "Taken from pGWiz"
FT   misc_feature   3415..3718
FT                  /*tag= i
FT                  /note= "Non-coding DNA residual from pNK2859"
FT   insertion_seq  3719..3761
FT                  /*tag= j
FT                  /note= "Left insertion sequence recognised by Tn10"
FT   misc_feature   3762..3831
FT                  /*tag= k
FT                  /note= "Non-coding lambda DNA residual from pNK2859"
FT   misc_feature   3832..3837
FT                  /*tag= l
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   misc_feature   3838..4044
FT                  /*tag= m
FT                  /note= "Multiple cloning site from pBluescript11 sk(-)"
FT   promoter       4049..5693
FT                  /*tag= n
FT                  /note= "CMV promoter/enhancer from pGWiz"
FT   misc_feature   5694..5701
FT                  /*tag= o
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   CDS            5702..6617
FT                  /*tag= p
FT                  /note= "DsRed reporter coding sequence including polyA
FT                  from pDSRed1.1"
FT   misc_feature   6618..7101
FT                  /*tag= q
FT                  /note= "Multiple cloning site from pBluescript11 sk(-)"
FT   misc_feature   7102..7106
FT                  /*tag= r
FT                  /note= "Residue from ligation of restriction enzyme sites
FT                  used in construction"
FT   insertion_seq  7107..7176
FT                  /*tag= s
FT                  /note= "Right insertion sequence recognised by Tn10"
FT   misc_feature   7177..7218
FT                  /*tag= t
FT                  /note= "Non-coding lambda DNA residual from pNK2859"
FT   misc_feature   7219..8062
FT                  /*tag= u
FT                  /note= "Non-coding DNA residual from pNK2859"
FT   misc_feature   8063..10263
FT                  /*tag= v
FT                  /note= "Remainder from pBluescript11 sk(-)"
XX
PN   WO200400315T-A2.
XX
PD   08-JAN-2004.
XX
PF   26-JUN-2003; 2003WO-US020389.
XX
PR   26-JUN-2002; 2002US-0392415P.
PR   21-JAN-2003; 2003US-0441377P.
PR   21-JAN-2003; 2003US-0441381P.
PR   21-JAN-2003; 2003US-0441392P.
PR   21-JAN-2003; 2003US-0441405P.
PR   21-JAN-2003; 2003US-0441447P.
PR   21-JAN-2003; 2003US-0441502P.
XX
PA   (TRAN-) TRANSGENRES LLC.
PA   (LOUU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.
XX
PI   Cooper RK, Cadd GG, Fioretti WC, Deboer KF;
XX
DR   WPI; 2004-083042/08.
XX
PT   New vector comprising a transposase gene operably linked to a promoter,
PT   useful in treating Alzheimer's disease, breast cancer, phenylketonuria,
PT   autoimmune diseases, bacterial or viral infections or Types I and II
PT   diabetes.
XX
PS   Example 2; SEQ ID NO 2; 150pp; English.
XX
CC   This sequence represents a new vector, pTnMod (CMV/Red), comprising a
CC   transposase gene operably linked to a first promoter and one or more gene
CC   of interest operably linked to one or more additional promoter, where
CC   the one or more genes of interest and their operably linked promoters are
CC   flanked by transposase insertion sequences recognized by the transposase
CC   and the first promoter comprises a modified Kozak sequence comprising
CC   ACCATG. The transposase is a Tn10 transposase and is modified in one to
CC   twenty of the first codons. The gene of interest in operably linked to a
CC   second promoter, which is a constitutive promoter or an inducible
CC   promoter, e.g. an ovalbumin or a vitellogenin promoter. The vector
```

CC   further comprises a polyA sequence, a conalbumin polyA sequence, operably
CC   linked to the transposase gene. The vector further comprises two stop
CC   codons operably linked to the transposase gene. The first or second gene
CC   of interest is operably linked to a second or third promoter,
CC   respectively. The first and a second gene of interest are operably linked
CC   to a second promoter. The vector further comprises an enhancer operably
CC   linked to the one or more genes of interest. The vector further comprises at
CC   least a portion of an ovalbumin enhancer. The enhancer comprises an
CC   egg directing sequence, ovalbumin or ovomucoid signal sequence or
CC   vitellogenin targeting sequence, operably linked to the one or more genes
CC   of interest. The vector is useful in producing transgenic animals with
CC   desired proteins or molecules. They are also useful in treating
CC   Huntington's disease, alpha-1-antitrypsin deficiency, Alzheimer's
CC   disease, breast cancer, cystic fibrosis, galactosaemia, congenital
CC   hypothyroidism, maple syrup urine disease, neurofibromatosis,
CC   phenylketonuria, sickle cell disease, Smith-Lemli-Optiz Syndrome,
CC   autoimmune diseases, shipping fever in cattle, mastitis, bacterial or
CC   viral infections or Types I and II diabetes.
XX
SQ   Sequence 10263 BP; 2648 A; 2595 C; 2374 G; 2646 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AD104102 (1-10263)

Alignment Scores:
Pred. No.:
Score: 3.31e-138
                       1210.00
Percent Similarity: 100.00%   Length: 10263
Best Local Similarity: 99.56%   Matches: 224
Query Match: 99.67%   Conservative: 1
DB: 12   Mismatches: 0
                                   Indels: 0
                                   Gaps: 0

Qy   1 MetArgSerSerLysSerAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||
Db 5725 GTGCGGCTCCTTCCAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 5784

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
        ||||||||||||||||||||||||||||||||||||||||||||
Db 5785 ACCGTGAACGGCCACGAGTTCGAAGCAACGGCCGAGGGCGAGGGCGATGCCCTACGAGGGC 5844

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
        ||||||||||||||||||||||||||||||||||||||||||||
Db 5845 CACAACACCCTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5904

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
        ||||||||||||||||||||||||||||||||||||||||||||
Db 5905 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5964

Qy   81 AspThrTyrLysLeuSerPheProGlyGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||
Db 5965 GACTACAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 6024

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||
Db 6025 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 6084

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||
Db 6085 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 6144

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgGlnGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||
Db 6145 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGGCTGGTGAAGGGCGAG 6204

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||
Db 6205 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 6264

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||
Db 6265 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 6324

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||
Db 6325 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 6384

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 6385 CACCACCTGTTCCTG 6399

RESULT 40
ABA95906
ID   ABA95906 standard; cDNA; 681 BP.
XX
AC   ABA95906;
XX
DT   29-MAY-2002 (first entry)
XX
DE   Yeast optimised RFP encoding cDNA SEQ ID NO 2.
XX
KW   Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW   Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW   Escherichia coli; green fluorescent protein; biotechnology; gene; ss.
XX
OS   Anthozoa.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..681
FT   /tag= a
FT   /product= "yeast optimised Red Fluorescent Protein"
XX
PA   (GPCB-) GPC BIOTECH AG.
XX
PR   27-JAN-2000; 2000DE-02001395.
XX
PF   27-JAN-2000; 2000DE-02001395.
PS   Claim 2; Fig 1; 19pp; German.
XX
PD   15-MAR-2001.
XX
PN   DE20001395-U1.
XX
DR   WPI; 2002-228394/29.
DR   P-PSDB; ABB08835.
XX
CC   The invention relates to DNA (II) containing either sequence ABA95905 or
CC   sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC   (yRFP). (I) are used to express red fluorescent protein (RFP) in
CC   eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC   plants, especially dicotyledonous plants including Nicotiana tabacum or
CC   Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC   especially Escherichia coli. RFP is useful in the same way as green
CC   fluorescent protein but is more generally applicable in modern
CC   biotechnology. (II) are optimised for expression in yeast and so generate
CC   RFP at higher levels with stronger fluorescence and thus lowers the
CC   detection limit and gives a better signal-to-noise ratio
XX
SQ   Sequence 681 BP; 198 A; 147 C; 161 G; 175 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ABA95906 (1-681)

Alignment Scores:
Pred. No.:
Score: 9.65e-140
                       1209.00
Percent Similarity: 100.00%   Length: 681
Best Local Similarity: 100.00%   Matches: 224
Query Match: 99.59%   Conservative: 0
DB: 6   Mismatches: 0
                                   Indels: 0
                                   Gaps: 0

Qy   2 ArgSerSerLysSerAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGlyThr 21
        |||||||||||||||||||||||||||||||||||||||||||||
Db   7 AGATCTTCTAAGAACGTCATCAAGGAATTCATGAGATTCAAGGTTAGAATGGAAGGTACT 66

```
Qy    22 ValAsnGlyHisGluPheGlnIleGluGlyGluGlyGluGlyArgProTyrGluGlyHis 41
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    67 GTTAACGGGTCACGAATTCGAAATCGAAAGGTGAAGGTGAAGGTAGACCATACGAAGGTCAC 126

Qy    42 AsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeu 61
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 AACACTGTCAAGTTGAAGGTTACTAAGGGTGGTCCATTGCCTTGGACATCTTG 186

Qy    62 SerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAsp 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 TCTCCACAATTCCAATATGGTTCTAAGGTCTACGTCAAGCACCCTGCCGATATTCCAGAC 246

Qy   102 GlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLys 121
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 GGTGGTGTTGCTACTGTTACTCAAGACTCCTCCTTGCAAGACGGTTGTTTCATCTACAAG 366

Qy    82 TyrLysLysLeuSerPheProGluGlyTyrGlyPheLysTrpGluArgValMetAsnPheGluAsp 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 TACAAGAAGTTGTCCTTCCCAGAAGGTTACGGTTTCAAGTGGGAAAGAGTTATGAACTTCGAAGAC 306

Qy   122 ValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThrMet 141
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 GTCAAGTTCATCGGTGTTGCTAACTTCCCATCTGACGGTCCAGTCATGCAAAAGAAGACTATG 426

Qy   142 GlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIle 161
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427 GGTTGGGAAGCTTCTACCGAACGTTTGTACCCAAGAGACGGTGTCTTGAAGGGTGAAATC 486

Qy   162 HisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyr 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   487 CACGGTGCCTTGAAAGTTGAAGGTGGTGGTCACTACTTGGTCGAATTCAAGTCTATCTAC 546

Qy   182 MetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIle 201
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   547 ATGGCTAAGAAGCCAGTCCAATTGCCAGGTTACTACTACGTTGACTCTAAGTTGGACATC 606

Qy   202 ThrSerHisAsnGluAspTyrThrIleValGluIleGlnTyrGluArgThrGluGlyArgHis 221
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   607 ACCTCTCACAACGAAGACTACACTATCGTCGAACAATACGAACGTACTGAAGGTAGACAC 666

Qy   222 HisLeuPheLeu 225
         ||||||||||
Db   667 CACTTGTTCTTG 678
```

```
Alignment Scores:
Pred. No.:
Score:                9.65e-140
Percent Similarity:   1209.00          Length:             681
Best Local Similarity: 100.00%          Matches:            224
Query Match:          100.00%          Conservative:       0
DB:                   99.59%           Mismatches:         0
                      6                Indels:             0
                                        Gaps:              0
                                        Other:             0
```

```
SQ    Sequence 681 BP; 198 A; 147 C; 161 G; 0 T; 175 U; 0 Other;
```

```
US-10-006-922A-12 (1-225) x ABA95907 (1-681)

Qy     2 ArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThr 21
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7 AGAUCUCUAAGAACGUUAUCAAGGAAUUCAUGAGAUUCAAGGUUAGAAUGGAAGGUACU 66

Qy    22 ValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHis 41
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    67 GUUAACGGUCACGAAUUCGAAAUCGAAGGUGAAGGUGAAGGUAGACCAUACGAAGGUCAC 126

Qy    42 AsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeu 61
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 AACACUGUCAAGUUGAAGGUUACUAAGGGUGGUCCAUUGCCAUGGGACAUCUUG 186

Qy    62 SerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAsp 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 UCUCCACAAUUCCAAUACGGUUCUAAGGUCUACGUCAAGCACCCAGCUGACAUUCCAGAC 246

Qy    82 TyrLysLysLeuSerPheProGluGlyTyrGlyPheLysTrpGluArgValMetAsnPheGluAsp 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 UACAAGAAGUUGUCCUUCCCAGAAGGUUACGGUUUCAAGUGGGAAAGAGUUAUGAACUUCGAAGAC 306

Qy   102 GlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLys 121
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 GGUGGUGUUGCUACUGUUACUCAAGACUCCUCCUUGCAAGACGGUUGUUUCAUCUACAAG 366

Qy   122 ValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThrMet 141
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 GUCAAGUUCAUCGGUGUUGCUAACUUCCCAUCUGACGGUCCAGUCAUGCAAAAGAAGACUAUG 426

Qy    42 AsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeu 61
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 AACACUGUCAAGUUGAAGGUUACUAAGGGUGGUCCAUUGCCAUGGGACAUCUUG 186

Qy    62 SerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAsp 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 UCUCCACAAUUCCAAUACGGUUCUAAGGUCUACGUCAAGCACCCAGCUGACAUUCCAGAC 246

Qy   142 GlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIle 161
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427 GGUUGGGAAGCUUCUACCGAACGUUUGUACCCAAGAGACGGUGUCUUGAAGGGUGAAAUC 486

Qy    82 TyrLysLysLeuSerPheProGluGlyTyrGlyPheLysTrpGluArgValMetAsnPheGluAsp 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 UACAAGAAGUUGUCCUUCCCAGAAGGUUACGGUUUCAAGUGGGAAAGAGUUAUGAACUUCGAAGAC 306

Qy   162 HisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyr 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   487 CACGGUGCCUUGAAAGUUGAAGGUGGUGGUCACUACUUGGUCGAAUUCAAGUCUAUCUAC 546

Qy   182 MetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIle 201
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   547 AUGGCUAAGAAGCCAGUCCAAUUGCCAGGUUACUACUACGUUGACUCUAAGUUGGACAUC 606

Qy   202 ThrSerHisAsnGluAspTyrThrIleValGluIleGlnTyrGluArgThrGluGlyArgHis 221
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   202 
```

```
Db
Qy  607 ACCUCUCACAACGAAGACUACACUAUCGUCGAACAUACGAACGUACUGAAGGUAGACAC 666
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Qy  222 HisLeuPheLeu 225
        ||||||||||||
Db  667 CACUUGUUCUG 678

Qy
Db

RESULT 42
ID   AAD28208 standard; cDNA; 678 BP.
XX
AC   AAD28208;
XX
DT   22-APR-2002 (first entry)
XX
DE   Discosoma sp. fluorescent protein E5 encoding cDNA.
XX
KW   Fluorescent timer protein; protein movement; translocation; trafficking;
KW   promoter activity; gene expression; transgenic plant; gene modification;
KW   protein age; E5; 88.
XX
OS   Discosoma sp.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..678
FT                  /tag= a
FT                  /product= "E5 protein"
XX
PN   WO200196373-A2.
XX
PD   20-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-US019097.
PR   14-JUN-2000; 2000US-0211607P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Fradkov AF, Terskikh A;
XX
DR   WPI; 2002-154595/20.
DR   P-PSDB; AAE17541.
XX
PT   New fluorescent timer proteins comprising an emission spectrum that
PT   changes over time from a first wavelength to a second wavelength, useful
PT   for monitoring intracellular protein movement, translocation, trafficking
PT   or stability.
XX
PS   Disclosure; Fig 2; 89pp; English.
XX
```

CC The invention relates to a fluorescent timer protein having an emission
CC spectrum that changes over time after synthesis from a first wavelength
CC to a second wavelength. The fluorescent timer proteins are useful in
CC monitoring the activity of a promoter, determining the age of a protein,
CC identifying an agent that modulates the activity of a promoter and in
CC enriching a population of cells comprising a fluorescent timer protein.
CC The fluorescent timer proteins are also useful for assessing gene
CC expression during development of a multicellular organism or during
CC cellular differentiation, in response to a drug or other inducer of
CC promoter activity, as a reporter to serve as a read-out of promoter
CC activity, monitoring intracellular protein movement or translocation,
CC protein trafficking, or protein stability, to investigate temporal
CC aspects of the activity of a regulatory element, for determining cell
CC fate during development and organ remodelling, in spatial and temporal
CC visualisation of newly synthesised proteins and accumulated proteins, and
CC in distinguishing between newly formed and pre-existing structures, e.g.
CC membrane junctions and extracellular matrix components. The fluorescent
CC timer proteins may further be used to investigations where photobleaching
CC techniques are employed, as detectable labels, as selectable markers, as
CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC assays, and as second messenger detectors. The nucleic acids can be used
CC to generate transgenic, non-human plants or animals or site-specific gene
CC modifications in cell lines. The present sequence is a cDNA encoding

CC Discosoma sp. E5 fluorescent protein derived from humanised wild-type
CC Anthozoa protein drFP583 by substituting Val to Ala at 105 and Ser to Thr
CC at 197
XX
SQ Sequence 678 BP; 147 A; 224 C; 203 G; 104 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:
Score:                  1207.00     Length:        678
Percent Similarity:     99.56%      Matches:       223
Best Local Similarity:  99.11%      Conservative:    1
Query Match:            99.42%      Mismatches:      1
DB:                     6           Indels:          0
                                    Gaps:            0

US-10-006-922A-12 (1-225) x AAD28208 (1-678)

Qy    1 MetArgSerSerLysAlaAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTGATGCAGAAGAAGACC 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGAAGTCCATC 540

Qy  181 TyrMetAlaLysLysProValGlyLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||:::
Db  541 TACATGGCCAAGAAGCCCGTGGGCCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  661 CACCACCTGTTCCTG 675

RESULT 43
ID   AAH47656 standard; cDNA; 681 BP.
XX
AC   AAH47656;
XX
DT   30-NOV-2001 (first entry)
XX
DE   Anthozoan red fluorescent protein synthetic mutant nucleotide sequence.
```

```
XX
KW   Fluorescent protein; Anthozoan; fluorescence; marker; FRET; mutant; ss.
OS   Synthetic.
XX
PN   WO20016291 9-A1.
XX
PD   30-AUG-2001.
XX
PF   13-FEB-2001; 2001WO-US004625.
PR   23-FEB-2000; 2000US-0184732P.
XX
PA   (AURO-) AURORA BIOSCIENCES CORP.
XX
PI   Nelson D, Zamaira E, Tsien R;
XX
DR   WPI; 2001-557704/62.
DR   P-PSDB; AAG65510.
XX
PT   Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
PT   functional red fluorescent proteins, and the encoding nucleic acids, with
PT   key mutations for improving the proteins function.
XX
PS   Disclosure; Page 86; 90pp; English.
XX
CC   The invention provides a nucleic acid encoding functional red fluorescent
CC   protein (II) that differs from the sequence of an Anthozoan red
CC   fluorescent protein by at least one amino acid substitution, and with
CC   different fluorescent properties. The red fluorescent protein of the
CC   invention can be expressed by standard recombinant methodology. (II) are
CC   used a fluorescent markers and FRET partners. It is used for identifying
CC   protein-protein interactions. (II) is also suitable for multiplexed
CC   fluorescent analysis and FRET-based applications using existing Aequorea
CC   fluorescent proteins. (II) has improved brightness, reduced spectral
CC   cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC   The key mutations in the encoding nucleic acids provide improved folding,
CC   brightness, and create (II) with sharper, more defined excitation and
CC   emission peaks when expressed in mammalian cells. The present sequence
CC   represents the nucleotide sequence of an improved synthetic mutant of an
CC   anthozoan fluorescent protein
XX
SQ   Sequence 681 BP; 160 A; 201 C; 220 G; 100 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:
Score:                3.02e-139
Score:                1205.00
Percent Similarity:   99.56%
Best Local Similarity: 99.11%
Query Match:          99.26%
DB:                   4
```

Matches:
```
Length:         681
Matches:        223
Conservative:   1
Mismatches:     1
Indels:         0
Gaps:           0
```

US-10-006-922A-12 (1-225) x AAH47656 (1-681)

```
Qy        1 MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly  20
            :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4 GTGAGGAGCAGCAAGAACGTGATCAAGGAGTTCATGAGGTTCAAGGTGCGCATGGAGGGC  63

Qy       21 ThrValAsnGlyHisGluPheIleGluGluGlyGluGlyGluGlyArgProTyrGluGly  40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       64 ACCGTGAACGGCCACGAGTTCACGAGGCAGGCGAGGGCGAGGGCCCTACGAGGGC  123

Qy       41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      124 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGGCCCCCCTTCGCCTGGGACATC  183

Qy       61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      184 CTGAGCCCCCCATTTCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  243

Qy       81 AspTyrLysLysLeuSerPheProGluGlyPhelysTrpGluArgValMetAsnPheGlu  100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      244 GACTACAAGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG  303
```

```
RESULT 44
ID   ADR30814 standard; DNA; 7508 BP.
AC   ADR30814;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Zebrafish BDNF gene construct DNA #2.
XX
KW   Brain-derived neurotrophic factor; BDNF; promoter; zebrafish; ds.
XX
OS   Danio rerio.
OS   Chimeric.
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FH
FT   misc_feature    1..20
FT                   /note= "Zebrafish BDNF gene 5' flank region"
FT                   /tag=  a
FT   promoter        1777..1804
FT                   /note= "Zebrafish BDNF promoter gene"
FT                   /tag=  d
FT   exon            1805..2099
FT                   /note= "Exon 1c (5' UT) of Zebrafish BDNF gene "
FT                   /tag=  e
FT   misc_feature    2100..2119
FT                   /note= "Vector DNA"
FT                   /tag=  f
FT   misc_feature    2120..2815
FT                   /note= "Vector DNA"
FT                   /tag=  c
FT   misc_feature    2816..2820
FT                   /note= "Reporter vector DNA"
FT                   /tag=  g
FT   misc_feature    2821..5099
FT                   /note= "Linker DNA"
FT                   /tag=  i
FT   misc_feature    21..1776
FT                   /note= "Expression vector DNA"
FT                   /tag=  b
```

```
Qy      101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      304 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCTGCTTCATCTAC  363

Qy      121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC  423

Qy      141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  483

Qy      161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  543

Qy      181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAspSerLeuAsp  200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGACTCCAAGCTGGAC  603

Qy      201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      604 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGAGGACCGAGGGCAGG  663

Qy      221 HisHisLeuPheLeu  225
            |||||||||||||||
Db      664 CACCACCTGTTCCTG  678
```

```
FT   misc_feature    5100..7508
FT                   /tag= 1
FT                   /note= "Vector DNA"
XX
PN   US20041572294-A1.
XX
PD   12-AUG-2004.
XX
PF   23-DEC-2003; 2003US-00742828.
XX
PR   30-NOV-2001; 2001US-0334079P.
PR   29-NOV-2002; 2002US-00306737.
XX
PA   (USGO ) US DEPT VETERANS AFFAIRS.
XX
PI   Heinrich G, Huynh G;
XX
DR   WPI; 2004-592772/57.
XX
PS   Claim 1; SEQ ID NO 4; 56pp; English.
XX
PT   Nucleic acid construct useful for screening substances for their ability
PT   to influence production of brain derived neurotrophic factor (BDNF),
PT   comprises promoter of BDNF gene and red fluorescent reporter.
XX
CC   The present invention provides an isolated brain-derived neurotrophic
CC   factor (BDNF) gene promoter and its associated nucleic acid constructs.
CC   The invention is useful for screening a number of substances for their
CC   ability to influence the production of BDNF in living organisms. The
CC   present sequence is useful for screening brain-derived neurotrophic factor (BDNF)
CC   gene construct DNA. This sequence is used in the invention.
XX
SQ   Sequence 7508 BP; 2249 A; 1604 C; 1665 G; 1990 T; 0 U; 0 Other;

US-10-006-922a-12 (1-225) x ADR30814 (1-7508)

Alignment Scores:
Pred. No.:
Score:                  1203.00
Percent Similarity:     99.56%          Length:         7508
Percent Identity:       98.67%          Matches:         222
Best Local Similarity:  99.09%          Conservative:      2
Query Match:            13              Mismatches:        1
                                        Indels:            0
                                        Gaps:              0

Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2123 GTGCGGCTCCTTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 2182

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2183 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2242

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2243 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2302

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2303 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2362

Qy    81 AspThrTyrLysLeuLysPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2363 GACTACAAGAGGGTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2422

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2423 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 2482

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2483 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 2542

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
```

```
Db  2543 ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2602
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   161 IleHisIleLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluGluLysSerIle 180
Db  2603 ATCCACACCAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGGAGAAGTCCATC 2662

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db  2663 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 2722

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2723 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2782

Qy   221 HisHisHisLeuPheLeu 225
         ||||||||||||||||
Db  2783 CACCACCTGTTCCTG 2797

RESULT 45
ACA55359
ID   ACA55359 standard; DNA; 6985 BP.
XX
AC   ACA55359;
XX
DT   06-JUN-2003 (first entry)
XX
DE   Transformation vector piggyBAC related plasmid p3xP3-DsRed-orf.
XX
KW   PiggyBac; transposon; eukaryotic transformation vector; ds;
KW   transformed cell; transformed embryo; transgenic; plasmid; cyclic;
KW   circular.
XX
OS   Synthetic.
XX
PN   US2002173634-A1.
XX
PD   21-NOV-2002.
XX
PF   30-OCT-2001; 2001US-00001189.
XX
PR   31-OCT-2000; 2000US-0244677P.
PR   01-NOV-2000; 2000US-0244984P.
XX
PA   (FRAS/) FRASER M J.
PA   (LIXX/) LI X.
PA   (BEAM/) BEAM T.
PA   (HUAV/) HUA-VAN A.
XX
PI   Fraser MJ, Li X, Beam T, Hua-Van A;
XX
DR   WPI; 2003-352597/33.
XX
PT   New DNA molecule in the transposon piggyBac, useful for transferring
PT   genes into host cells or embryos for transforming the cells of embryos
PT   that can be used in making transgenic organisms.
XX
PS   Example 10; Fig 9(B); 151pp; English.
XX
CC   The invention describes a DNA molecule comprising at least 163
CC   consecutive nucleotide base pairs of the 3' terminal region beginning at
CC   the 3' terminal base pair, and at least 125 consecutive nucleotide base
CC   pairs of the 5' terminal region beginning at the 5' terminal base pair, of
CC   the piggyBac molecule. The region extends from the restriction site SacI
CC   to the end of the piggyBac molecule. The DNA molecule in the transposon
CC   piggyBac is useful for transferring genes into host cells or embryos for
CC   transforming the cells of embryos. The transformed cells or embryos are
CC   useful for developing or making transgenic organisms. This sequence
CC   represents plasmid used in the creation of minimal sequence eukaryotic
CC   transformation vector piggyBac
XX
SQ   Sequence 6985 BP; 1860 A; 1680 C; 1731 G; 1714 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:
Score: 7.86e-137
Percent Similarity: 1197.00
Best Local Similarity: 99.56%
Query Match: 99.12%
DB: 98.60%

Length: 6985
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 1
Gaps: 0

US-10-006-922A-12 (1-225) x ACA55359 (1-6985)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ::::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2974 GTGCGCGCTTCCTTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3033

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3034 ACGGTGAACGGCCACGAGTTCGAGATCGAAGGGGAAGGCGAGGGCCGCCCCTACGAGGGC 3093

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3094 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3153

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3154 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3213

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3214 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3273

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3274 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3333

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3334 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 3393

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3394 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3453

Qy 161 IleHisGlys-AlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIl 180
        ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db 3454 ATCCACAAGGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCAT 3513

Qy 180 eTyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAs 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3514 CTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGAA 3573

Qy 200 pIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyAr 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3574 CATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCG 3633

Qy 220 gHisHisHisLeuPheLeu 225
        |||||||||||||||||||
Db 3634 CCACCACCTGTTCCTG 3649

RESULT 46
ABL61142
ID   ABL61142 standard; DNA; 666 BP.
XX
AC   ABL61142;
XX
DT   18-SEP-2002   (first entry)
XX
DE   Red fluorescent protein reporter DNA fragment.
XX
KW   Alpha-SMA; smooth muscle actin; screening; smooth muscle cell;
KW   myofibroblast; gene therapy; red fluorescent protein; ds.
XX
OS   Synthetic.
XX

PN   EP1172375-A1.
XX
PD   16-JAN-2002.
XX
PF   22-DEC-2000; 2000EP-00128446.
XX
PR   11-JUL-2000; 2000DE-010336633.
PR   31-OCT-2000; 2000DE-01053879.
XX
PA   (ODEN/) ODENTHAL M.
XX
PI   Odenthal M, Jung D;
XX
DR   WPI; 2002-149590/20.
XX
PT   New nucleic acid containing regulatory region of the smooth muscle actin
PT   gene, useful e.g. for manipulating gene expression in smooth muscle
PT   cells.
XX

Disclosure; Page 20; 44pp; German.

PS   Disclosure; Page 20; German.
XX
CC   This invention describes a novel nucleic acid (I) comprising: (i) at
CC   least one functional region (Ia) from the regulatory region of the alpha-
CC   smooth muscle actin (SMA) gene and (ii) at least one additional
CC   functional sequence (Ib) operably linked to (Ia). The products of the
CC   invention can be used for preparing genetically modified eukaryotic cells
CC   or organisms, for isolation and screening of smooth muscle cells,
CC   myofibroblasts or related cells, and for manipulation of gene expression
CC   and/or cell function in smooth muscle cell or myofibroblasts,
CC   particularly for gene therapy. Component (Ia) provides cell-type- or
CC   differentiation-specific expression or modulation of genes. This sequence
CC   represents a DNA fragment of red fluorescent protein (EGFP) which can be
CC   used as a reporter molecule under the control of the alpha-SMA described
CC   in the invention
XX
SQ   Sequence 666 BP; 202 A; 126 C; 176 G; 162 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 3.82e-138
Percent Similarity: 1196.00
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 98.52%

Length: 666
Matches: 221
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x ABL61142 (1-666)

Qy   5 LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly 24
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGGATGGAAGGAACGGTCAATGGG 60

Qy   25 HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal 44
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CACGAGTTTGAAATAGAAGGGGAAGGAGAGGGCCCATACGAAGGCCACAATACCGTA 120

Qy   45 LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAGCTTAAGGTAACCAAGGGGGGACCTTTGCCCATTTGCTTGGGGATTATTTTGTCACCACAA 180

Qy   65 PheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLys 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCAGACTATAAAAAG 240

Qy   85 LeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyVal 104
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTATGAACTTTGAAGACGGTGGCGTC 300

Qy 105 ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe 124
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTACAAGGTTCAAGTTC 360

Qy 125 IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu 144
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 361 ATGGGCGTGAACTTTCCTTCCGGATGGAACCTGTTATGCAAAAGAGAATGGGCGTGGGAA 420
Qy |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 145 AlaSerThrGluArgHisLeuTyrProArgAspGlyValLeuLysGlyGlyHisGlyHisGlyAla 164

Db 421 GCCAGCACTGAGCCGTTTGTATCCTCGTGATGGCTGTTGAAAGGAGAGATTCATAAGGCT 480
Qy |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 165 LeuGluLysGlyAspGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys 184

Db 481 CTGAAGCTGAAAGACGGGTGGTCATTACCTAGTTGAATTCAAAAAGTATTTACATGGCAAAG 540
Qy |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 185 LysProValGlnLeuProGlyTyrTyrTyrValAspSerIleLysLeuAspIleThrSerHis 204

Db 541 AAGCCTGTGCAGCTACCAGGTACTACTATGTTGACTCAAAACTGGATATAACAAGCCAC 600
Qy |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 205 AsnGluAspTyrThrIleValGluGlnIleTyrGluArgThrGluGlyGlyHisHisLeuPhe 224

Db 601 AACGAAGACTATACAATCGTTGAGCAGATTTATGAAAGAACCGAGGGAGGACGCCACCATCTGTTC 660
Qy |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 661 CTT 663
Qy |||
Qy 225 Leu 225

RESULT 47

ID   ADO78067 standard; DNA; 711 BP.
XX
AC   ADO78067;
XX
DT   26-AUG-2004 (first entry)
XX
DE   Corallimorpharia red fluorescent protein mutant RedI DNA.
XX
OS   Corallimorpharia.
XX
FH   Key   Location/Qualifiers
FT
FT   CDS   1..711
FT   /tag= a
FT   /product= "Corallimorpharia red fluorescent protein
FT   mutant RedI"
XX
PN   US2004110225-A1.
XX
PD   10-JUN-2004.
XX
PF   09-DEC-2002; 2002US-00314936.
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2004-467659/44.
DR   P-PSDB; ADO78068.
XX
PT   New first mutant fluorescent protein having an optical property
PT   relatively different to a corresponding optical property produced by a
PT   wild-type red or green fluorescent protein, useful for detecting
PT   expression of a gene.
XX
PS   Claim 9; SEQ ID NO 1; 30pp; English.
XX
CC   The invention describes a first mutant fluorescent protein (I) comprising
CC   an amino acid sequence that differs from that of a corresponding wild-
CC   type red fluorescent protein by an amino acid substitution, where the

first mutation fluorescent protein mutant has an optical property that
differs relative to the corresponding optical property produced by the
wild-type red or green fluorescent protein. (I) is useful for detecting
expression of gene. The polynucleotide (II) encoding (I) is useful for
detecting expression of a gene which involves introducing (II) into a
cell or organism, allowing the cell to replicate, and detecting
expression of the nucleic acid expression is detected in vivo or in
vitro. (III) is useful for regenerating an animal such as a zebrafish
which contains (II). (II). (I) is also useful for detecting expression
of a gene, in biochemical assays, and as reagents. (II) is useful in
multiple labeling systems, as in vivo markers such as in mRNA
microinjection assays, and as taxonomic markers for studies of genetics,
colour indicators in diagnostic kits, coloured food additives, and
cosmetics ingredients. (I) is useful in research for up or down
regulation, to monitor promoter activity, to allow longer term monitoring
and to localise proteins. (I) has enhanced properties such as
substantially enhanced fluorescence and reduced toxicity. This sequence
encodes RedI, a mutant of the red fluorescent protein isolated from an
aquatic species believed to be either an Actinodiscus or Discosoma
species.

XX
SQ   Sequence 711 BP; 214 A; 131 C; 189 G; 177 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADO78067 (1-711)

Alignment Scores:
Pred. No.:   4.18e-138
Score: 1196.00
Percent Similarity: 99.11%   Matches: 221
Best Local Similarity: 98.22%   Conservative: 2
Query Match: 98.52%   Mismatches: 2
DB: 12   Indels: 0
                                   Gaps: 0

Qy   1 MetArgSerLysAspAsnValIleLysGluGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGTTGTTCCAGAAATGTTATCAAAGGAGTTCATGAGGTTTAGGTTCGTATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuSerPheProGluGlyPheAsnTrpGluArgValMetAsnPheGlu 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATAACCGTAAAGCTTAAGGTAACCAAGGAGGGGAACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACATCCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyGlyProValMetGlnLysLysThr 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGAGATTAAAATGGGAAAAGGGGTCAGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgHisLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCCGTTTGTATCCTCGTGATGGCGTTGAAAGGAGAG 480

Qy 161 IleHisLysGlyAlaLeuLysLysGlyAspGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||::|||
Db 481 ATTCATAAGGCTCTGAAGTTGAAAGACGGGTGGTCATTACCTAGTTGAATTCAAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCCTGTCGCGCTACCAGGGTACTACTATGTTGACTCAAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluArgThrGluGlyArg 220
        ||||||||||||||:::|||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACAAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCACTCGTTCCTT 675

CC Corallimorpharia red fluorescent protein mutant RedII DNA.
XX
DE Corallimorpharia red fluorescent protein mutant RedII.
XX
OS Corallimorpharia.
XX
FH Key   Location/Qualifiers
FT CDS   1..711
FT   /*tag= a
FT   /product= "Corallimorpharia red fluorescent protein
FT   mutant RedII"
XX

CC fluorescent protein; red fluorescent protein; green fluorescent protein;
KW gene expression detection; regeneration; multiple labeling system;
KW in vivo marker; microinjection assay; taxonomic marker; colour indicator;
KW food additive; cosmetic; Actinodiscus; Discosoma; gene; dsi; mutant;
KW RedII; coral-like anemone.
XX
DT 26-AUG-2004 (first entry)
XX
AC AD078069;
XX
ID AD078069 standard; DNA; 711 BP.
RESULT 48
AD078069
Db 661 CACCACTCGTTCCTT 675

Qy 221 HisHisLeuPheLeu 225

Qy 601 ATAACAAGCCACAACAAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluArgThrGluGlyArg 220

CC species.
CC aquatic species believed to be either an Actinodiscus or Discosoma
CC encodes RedII, a mutant of the red fluorescent protein isolated from an
CC substantially enhanced fluorescence and reduced toxicity. This sequence
CC and to localise proteins. (I) has enhanced properties such as
CC regulation, to monitor promoter activity, to allow longer term monitoring
CC cosmetics ingredients. (I) is useful in research for up or down
CC colour indicators in diagnostic kits, coloured food additives, and

SQ Sequence 711 BP; 215 A; 131 C; 188 G; 177 T; 0 U; 0 Other;
XX
DB:
Best Local Similarity: 98.52%   Gaps: 0
Query Match: 12   0
Percent Similarity: 99.11%   Conservative: 2
                                    Mismatches: 2
Score: 1196.00   Matches: 221
Pred. No.:   4.18e-138 Length: 711
Alignment Scores:

US-10-006-922A-12 (1-225) x AD078069 (1-711)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGTTGTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGCCACGAGTTTGAAATAGAAGGCGAAGGAGAAGGCCAATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAAGGCGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTTTATGTCAAGCATCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTCCCTGAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGCGCGTTCGTTACTGTAACCCAGGATTCCAGTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTTGTTTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGTTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAACTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluArgThrGluGlyArg 220
        ||||||||||||||:::|||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACAAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCACTCGTTCCTT 675

RESULT 49
AD284222
ID AD284222 standard; DNA; 711 BP.
XX

PN US2004110225-A1.
PD 10-JUN-2004.
XX
PR 09-DEC-2002; 2002US-00314936.
PF 09-DEC-2002; 2002US-00314936.
XX
PA (GIBB/) GIBBS P D L.
PA (CART/) CARTER R W.
PA (SCHM/) SCHMALE M C.
XX
PI Gibbs PDL, Carter RW, Schmale MC;
XX
DR WPI; 2004-467659/44.
DR P-PSDB; AD078070.
XX
PT New first mutant fluorescent protein having an optical property
PT relatively different to a corresponding optical property produced by a
PT wild-type red or green fluorescent protein, useful for detecting
PT expression of a gene.
XX
PS Claim 10; SEQ ID NO 3; 30pp; English.

The invention describes a first mutant fluorescent protein (I) comprising
an amino acid sequence that differs from that of a corresponding wild-
type red fluorescent protein by an amino acid substitution, where the
first mutation fluorescent protein mutant has an optical property that
differs relative to the corresponding optical property produced by the
wild-type red or green fluorescent protein. (I) is useful for detecting
expression of gene. The polynucleotide (II) encoding (I) is useful for
detecting expression of a gene which involves introducing (II) into a
cell or organism, allowing the cell to replicate, and detecting
expression of the nucleic acid by emission of fluorescent light. The
expression of the nucleic acid expression is detected in vivo or in
vitro. (II) is useful as a marker for detecting expression
of a gene, in biochemical assays, and as reagents. (I) is also useful in
multiple labeling systems, as in vivo markers such as in mRNA
microinjection assays, and as taxonomic markers for studies of genetics,
which contains (II). (II) is useful for regenerating an animal such as a zebrafish

```
AC   AD284222;
XX
DT   14-JUL-2005 (first entry)
XX
DE   Actinodiscus/Discosoma RFP mutant Red I-encoding DNA, SEQ ID NO:1.
XX
KW   Red fluorescent protein; Red I; mutant; gene; ds.
XX
OS   Corallimorpharia.
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            1..711
FT                  /*tag= a
FT                  /product= "Mutant Ac/DsRFP Red I"
FT   mutation       replace(694,.A)
FT                  /*tag= b
FT                  /note= "This nucleotide is A in the Red II mutant,
FT                  resulting in Asn rather than Asp at position 232 of the
FT                  protein"
XX
PN   US2005100954-A1.
XX
PD   12-MAY-2005.
XX
PF   23-DEC-2004; 2004US-00021014.
XX
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2005-343599/35.
DR   P-PSDB; AD284223.
XX
PT   Novel mutant fluorescent protein comprising amino acid sequence that
PT   differs from wild-type red or green fluorescent protein by amino acid
PT   substitutions, useful as markers for detecting desired gene expression.
XX
PS   Claim 3; SEQ ID NO 1; 29pp; English.
XX
CC   The invention relates to mutant red and green fluorescent proteins which
CC   have higher fluorescence intensities compared to the wild-type proteins
CC   from which they are derived. The invention also discloses nucleic acids
CC   encoding the mutant fluorescent proteins, and transgenic animals
CC   comprising such a nucleic acid. The mutant red fluorescent proteins (RFP)
CC   of the invention, Red I (AD284223) and Red II (AD284225), are derived
CC   from a novel wild-type RFP (referred to as Ac/DsRFP) isolated from a
CC   mushroom coral believed to be either an Actinodiscus or Discosoma
CC   species, while the mutant green fluorescent proteins (GFP), Green I
CC   (AD284227) and Green (AD284229), are derived from a novel wild-type GFP
CC   (known as McGFP) isolated from the great star coral Montastraea
CC   cavernosa. The Ac/DsRFP and McGFP mutants were obtained by subjecting the
CC   wild-type Ac/DsRFP and McGFP cDNAs to low-stringency PCR to introduce
CC   random mutations, cloning the PCR products into bacterial expression
CC   vectors, and selecting the transformed bacteria having higher relative
CC   fluorescence compared to those expressing the wild-type proteins. The
CC   mutant fluorescent proteins of the invention may be used in a wide range
CC   of applications. They are useful as markers for detecting the expression
CC   of a gene of interest, or can be used as in vivo markers in mRNA
CC   microinjection assays in transgenic animals. They can be used in
CC   expression systems (e.g., as components of fusion proteins), in multiple
CC   labeling systems, or can be used in biochemical assays and as reagents.
CC   They may additionally be used as taxonomic markers for studies of
CC   cnidarian genetics, as color indicators in diagnostic kits, as colored
CC   food additives and as cosmetic ingredients. The present sequence
CC   represents DNA encoding the mutant Actinodiscus/Discosoma red fluorescent
CC   protein Red I.
XX
SQ   Sequence 711 BP; 214 A; 131 C; 189 G; 177 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              4.18e-138
Score:                  1196.00          Length:         711
Percent Similarity:     99.11%           Matches:        221
Best Local Similarity:  98.22%           Conservative:   2
Query Match:            98.52%           Mismatches:     2
                        14               Indels:         0
DB:                                      Gaps:           0
```

US-10-006-922A-12 (1-225) x AD284222 (1-711)

```
Qy          1 MetArgSerLeuLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTTGTTCCAAGAATGTTATCAAGAAGTTCATGAGGTTCAAGGTTCGTATGGAAGGA 60

Qy         21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACCGAAGGC 120

Qy         41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIlePro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACAATACCGTAAAGCTTAAAGTTACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy         61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGTCACCACAATTTCAGTATGGGAGCAAGGTATATGTCAAGCATCCTGCCGACATACCA 240

Qy         81 AspThrTyrLysLeuLysSerProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACTATAAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy        101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACGGTGGCGTCGTTACTGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy        121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA 420

Qy        141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGGGGCTGGAGGCAAGCCCACACTGAGCGTTTGTATCCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy        161 IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluValLysSerIle 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||
Db        481 ATTCATCACAAGGCTCTCAAGTTGAAAGACGGTGGTCATTACCTAGTTGAAGTTAAAACTATT 540

Qy        181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy        201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuTyrGluArgThrGluGlyArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATAACAAGCCACAACAAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy        221 HisHisHisLeuPheLeu 225
              ||||||||||||||||
Db        661 CACCATCTGTTCCTT 675
```

```
RESULT 50
AD284224
ID   AD284224 standard; DNA; 711 BP.
XX
AC   AD284224;
XX
DT   14-JUL-2005 (first entry)
XX
DE   Actinodiscus/Discosoma RFP mutant Red II-encoding DNA, SEQ ID NO:3.
XX
KW   Red fluorescent protein; Red II; mutant; gene; ds.
XX
OS   Corallimorpharia.
OS   Synthetic.
```

```
XX
FH   Key            Location/Qualifiers
FT   CDS            1..711
FT                  /tag= a
FT                  /product= "Mutant Ac/DsRFP Red II"
XX
FT   mutation       replace(694,G)
FT                  /tag= b
FT                  /note= "This nucleotide is G in the Red I mutant,
FT                  resulting in Asp rather than Asn at position 232 of the
FT                  protein"
XX
PN   US2005100954-A1.
XX
PD   12-MAY-2005.
XX
PF   23-DEC-2004; 2004US-00021014.
XX
PR   09-DEC-2002; 2002US-00314936.
XX
PA   (GIBB/) GIBBS P D L.
PA   (CART/) CARTER R W.
PA   (SCHM/) SCHMALE M C.
XX
PI   Gibbs PDL, Carter RW, Schmale MC;
XX
DR   WPI; 2005-345399/35.
DR   P-PSDB; AD284225.
XX
PT   Novel mutant fluorescent protein comprising amino acid sequence that
PT   differs from wild-type red or green fluorescent protein by amino acid
PT   substitutions, useful as markers for detecting desired gene expression.
XX
PS   Example 4; SEQ ID NO 3; 29pp; English.
XX
CC   The invention relates to mutant red and green fluorescent proteins which
CC   have higher fluorescent intensities compared to the wild-type proteins
CC   from which they are derived. The invention also discloses nucleic acids
CC   encoding the mutant fluorescent proteins, and transgenic animals
CC   comprising such a nucleic acid. The mutant red fluorescent proteins (RFP)
CC   of the invention, Red I (AD284223) and Red II (AD284225), are derived
CC   from a novel wild-type RFP (referred to as Ac/DsRFP) isolated from a
CC   mushroom coral believed to be either an Actinodiscus or Discosoma
CC   species, while the mutant green fluorescent proteins (GFP), Green I
CC   (AD284227) and Green (AD284229), are derived from a novel wild-type GFP
CC   (known as McGFP) isolated from the great star coral Montastraea
CC   cavernosa. The Ac/DsRFP and McGFP mutants were obtained by subjecting the
CC   wild-type Ac/DsRFP and McGFP cDNAs to low-stringency PCR to introduce
CC   random mutations, cloning the PCR products into bacterial expression
CC   vectors, and selecting the transformed bacteria having higher relative
CC   fluorescence compared to those expressing the wild-type proteins. The
CC   mutant fluorescent proteins of the invention may be used in a wide range
CC   of applications. They are useful as markers for detecting the expression
CC   of a gene of interest, or can be used as in vivo markers in mRNA
CC   microinjection assays in transgenic animals. They can be used in
CC   expression systems (e.g., as components of fusion proteins), in multiple
CC   labeling systems, or can be used in biochemical assays and as reagents.
CC   They may additionally be used as taxonomic markers for studies of
CC   cnidarian genetics, as color indicators in diagnostic kits, as colored
CC   food additives and as cosmetic ingredients. The present sequence
CC   represents DNA encoding the mutant Actinodiscus/Discosoma red fluorescent
CC   protein Red II. Red II was found to have a fluorescent intensity at least
CC   50% greater than the mutant RFP Red I (AD284223).
XX
SQ   Sequence 711 BP; 215 A; 131 C; 188 G; 177 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:				
Pred.:	4.18e-138		Length:	711
Score:	1196.00		Matches:	221
Percent Similarity:	99.11%		Conservative:	2
Best Local Similarity:	98.22%		Mismatches:	2
Query Match:	98.52%		Indels:	0
DB:	14		Gaps:	0

```
US-10-006-922A-12 (1-225) x AD284224 (1-711)

Qy    1  MetArgSerLysAsnValIleLysGlyGluPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTTGTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheIleIleGluGlyGluGlyLysGlyLysProTyrGluGly  40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGGCCAATACGAAGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCGGCCGACATACCA  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGCGGTGGCCGTCGTTACTGTAACCCAGGATTCCAGTTTCAGGATGGCTGTTTCATCTAC 360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||::||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTCTTGAAAGGAGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||::|||
Db  481  ATTCATAAGGCTCTGAAGTTGAAAGACGGTGGTCACTACTTAGTGGAATTCAAAAGTATT  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluIleGlnGlnHisTyrThrGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db  601  ATAACAAGCCACAACGAAGACTATACAATTGTTGAGCAGCATTATACAGGGAGGC  660

Qy  221  HisHisIleLeuPheLeu 225
         |||||||||||||||
Db  661  CACCATCTGTTCCTT 675
```

RESULT 51
AAD46282

```
ID   AAD46282 standard; DNA; 678 BP.
XX
AC   AAD46282;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant DNA, E5-NA.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; gene.
KW   mutant; ds.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key              Location/Qualifiers
FT   CDS              1..678
FT                    /tag= a
```

FT   /product= "drFP583 mutant protein, E5-NA"
XX
PN   WO20026459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270908P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   P-PSDB; AAE28837.
XX
PT   New nucleic acid encoding a non-aggregating Cnidarian chromo- or
     fluorescent protein or mutant for
     analyte detection assays or fluorescence activated cell sorting
     applications.
PS   Disclosure; Page 73; 80pp; English.
CC   The invention relates to nucleic acid molecules encoding non-aggregating
     chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
     useful in analyte detection assays, as colouring agents, as markers in
     recombinant DNA applications, as subscreens or filters, in fluorescence
     resonance energy transfer (FRET) applications, as biosensors in
     prokaryotic and eukaryotic cells, in screening assays, as second
     messenger detectors, in fluorescence activated cell sorting applications,
     in protease cleavage assays or as fluorescent timers. The present
     sequence is Discosoma sp. drFP583 (NFP-6) mutant DNA of the invention
SQ   Sequence 678 BP; 145 A; 225 C; 203 G; 105 T; 0 U; 0 Other;

Alignment Scores:
Pred. No. :
Score: 1191.00   Length: 678
Percent Similarity: 98.67%   Matches: 220
Best Local Similarity: 97.78%   Conservative: 2
Query Match: 98.11%   Mismatches: 3
DB: 6   Indels: 0
                                      Gaps: 0

US-10-006-922A-12 (1-225) x AAD46282 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGlyGly 20
         ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCGTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGGATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGCCGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCAGATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 52
AAD28209
ID   AAD28209 standard; cDNA; 678 BP.
XX
AC   AAD28209;
XX
DT   22-APR-2002 (first entry)
XX
DE   Discosoma sp. non-aggregating mutant timer protein B5NA encoding cDNA.
XX
KW   Fluorescent timer protein; protein movement; translocation; trafficking;
     promoter activity; gene expression; transgenic plant; gene modification;
     protein age; mutant; B5NA; ss.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /tag= a
FT   /product= "Mutant B5NA protein"
XX
PN   WO20019373-A2.
XX
PD   20-DEC-2001.
XX
PF   14-JUN-2000; 2000US-0211607P.
XX
PF   13-JUN-2001; 2001WO-US019097.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Fradkov AF, Terskikh A;
XX
DR   WPI; 2002-154595/20.
DR   P-PSDB; AAE17542.
XX
PT   New fluorescent timer proteins comprising an emission spectrum that
     changes over time from a first wavelength to a second wavelength, useful
     for monitoring intracellular protein movement, translocation, trafficking
     or stability.
PS   Disclosure; Fig 3; 89pp; English.
CC   The invention relates to a fluorescent timer protein having an emission
     spectrum that changes over time after synthesis from a first wavelength
     to a second wavelength. The fluorescent timer proteins are useful in
     monitoring the activity of a promoter, determining the age of a protein,
     identifying an agent that modulates the activity of a promoter and in
     enriching a population of cells comprising a fluorescent timer protein.
     The fluorescent timer proteins are also useful for assessing gene

Db
Qy 361 AAGGTGGAAGTTCATCGGGGTGGAACTTCCCCGACGGCGCCCGTGATGCAGAGAAGACC 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 221 HisHisLeuPheLeu 225
         |||||||||||||||
Db 661 CACCACCTGTTCCTG 675

CC expression during development of a multicellular organism or during
CC cellular differentiation, in response to a drug or other inducer of
CC promoter activity, as a reporter to serve as a read-out of promoter
CC activity, monitoring intracellular protein movement or translocation,
CC protein trafficking, or protein stability, to investigate temporal
CC aspects of the activity of a regulatory element, for determining cell
CC fate during development and organ remodelling, in spatial and temporal
CC visualisation of newly synthesised proteins and accumulated proteins, and
CC in distinguishing between newly formed and pre-existing structures, e.g.
CC membrane junctions and extracellular matrix components. The fluorescent
CC timer proteins may further be used to investigations where photobleaching
CC techniques are employed, as detectable labels, as selectable markers, as
CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC assays, and as second messenger detectors. The nucleic acids can be used
CC to generate transgenic, non-human plants or animals or site-specific gene
CC modifications in cell lines. The present sequence is a cDNA encoding
CC Discosoma sp. E5NA non-aggregating fluorescent timer protein derived from
CC fluorescent timer protein E5 by substituting amino acids at positions
CC R2A, K5E and K9T. E5 protein is derived from humanised wild-type Anthozoa
CC protein drFP583 by substituting Val to Ala at 105 and Ser to Thr at 197
XX
SQ Sequence 678 BP; 145 A; 226 C; 203 G; 104 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 1.63e-137
Score: 1191.00
Percent Similarity: 98.67%   Length: 678
Best Local Similarity: 97.78%   Matches: 220
Query Match: 98.11%   Conservative: 2
DB: Mismatches: 3
                                 Indels: 0
                                 Gaps: 0

US-10-006-922A-12 (1-225) x AAD28209 (1-678)

```
Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
         ||  ||||||||||| ||||||||||||||||
Db   1 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy  81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACACCTACAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy 161 IleHisHisAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCACCACGCCCTGAAGCTGAAGGACGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 221 HisHisHisPheLeu 225
Db 661 CACCACCACTTCCTG 675
```

RESULT 53
AAD46281
ID   AAD46281 standard; DNA; 675 BP.
XX
AC   AAD46281;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583 (NFP-6) mutant DNA, E57-NA.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; NFP-6; gene;
KW   mutant; ds.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..675
FT   /*tag= a
FT   /product= "drFP583 mutant protein, E57-NA"
FT   /note= "No stop codon"
FT   /partial
XX
PN   WO20020268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-02709983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   P-PSDB; AAD28836.
XX
PT   New nucleic acid encoding a non-aggregating Cnidarian chromo- or fluorescent mutant
PT   of an aggregating chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page 73; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, in screening assays, as
CC   prokaryotic and eukaryotic cells, in fluorescence activated cell sorting
CC   messenger detectors, in fluorescence assays, as biosensors, as second
CC   in protease cleavage assays or as fluorescence activated timers. The present
CC   sequence is Discosoma sp. drFP583 (NFP-6) mutant DNA of the invention
XX
SQ   Sequence 675 BP; 142 A; 227 C; 204 G; 102 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 6.74e-137   Length: 675
                    1186.00   Matches: 219

```
Percent Similarity:    98.22%              Conservative:   2
Best Local Similarity: 97.33%              Mismatches:     4
Query Match:           97.69%              Indels:         0
DB:                    6                   Gaps:           0

US-10-006-922A-12 (1-225) x AAD46281 (1-675)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGgccTCCTCCGGAAGCGTTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 ACCGTGAACGGCCACCAGGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy  81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAGCCTGTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 221 HisHisLeuPheLeu 225
         |||||||||||||||
Db 661 CACCACCTGTTCCTG 675
```

RESULT 54
ADV96699

```
ID   ADV96699 standard; DNA; 747 BP.
XX
AC   ADV96699;
XX
DT   10-MAR-2005 (first entry)
XX
DE   Plant cell gene silencing related dsRed gene coding region.
XX
KW   gene silencing; plant; short interfering RNA; siRNA; RNA interference;
KW   phage resistance; ds; gene; dsRed.
XX
OS   Synthetic.
XX
PN   US2004261149-A1.
XX
PD   23-DEC-2004.
```

```
XX
PF   24-FEB-2004; 2004US-00785862.
XX
PR   24-FEB-2003; 2003US-0449546P.
XX
PA   (FAUQ/) FAUQUET C M.
PA   (PADM/) PADMANABHAN C.
PA   (RAMA/) RAMACHANDRAN V.
XX
PI   Fauquet CM, Padmanabhan C, Ramachandran V;
XX
PR   WPI; 2005-038811/04.
XX
PT   Suppressing germiniviral genes in a plant cell comprises introducing a
PT   short interfering RNA into the plant cell.
XX
PS   Example 1; SEQ ID NO 10; 15pp; English.
XX
CC   The invention relates to a novel method for suppressing the expression of
CC   a target gene in a plant cell, comprising introducing a short interfering
CC   RNA (siRNA) into the plant cell. The method for suppressing the
CC   expression of a target gene in a plant cell comprises: introducing the
CC   siRNA into the plant cell, where the siRNA is a double-stranded molecule
CC   with a first strand consisting of a nucleic acid sequence which is
CC   substantially complementary to a nucleic acid sequence of a target gene
CC   and a second strand consisting of a nucleic acid sequence which is
CC   substantially complementary to the first strand, where the siRNA is
CC   capable of suppressing the expression of the target gene. The method,
CC   siRNA and composition are useful for suppressing expression of a target
CC   gene in a plant cell. Particularly, they are useful for suppressing or
CC   inhibiting viral gene expression, in plant cells. In a particular
CC   expression, in plant cells. In a particular embodiment of the invention,
CC   siRNAs are targeted to the coding region (Rep) of the replication-
CC   associated protein (Rep) of an ACMV-CM strain. This polynucleotide
CC   sequence represents the coding region of a gene used in the plant gene
CC   suppression method of the invention.
XX
SQ   Sequence 747 BP; 159 A; 248 C; 224 G; 116 T; 0 U; 0 Other;
XX
```

```
Alignment Scores:
Pred. No.:
Score:                 1186.00            Length:         747
                                          Matches:        219
Percent Similarity:    98.22%             Conservative:   2
Best Local Similarity: 97.33%             Mismatches:     4
Query Match:           97.69%             Indels:         0
DB:                    14                 Gaps:           0

US-10-006-922A-12 (1-225) x ADV96699 (1-747)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGgccTCCTCCGGAAGCGTTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 ACCGTGAACGGCCACCAGGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy  81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAGCCTGTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360
```

Qy 121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysGlyThr 140
Db 361 AAGGTGAAGTTCATCGGCGTGGCGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACCTGTTCCTG 675

RESULT 55
ID   ADW43589 standard; DNA; 4200 BP.
XX
AC   ADW43589;
XX
DT   07-APR-2005 (first entry)
XX
DE   UbCP-lox-mirMSOD2-EGFP-lox-RFP.
XX
KW   ss; hairpin; Neuroprotective; shRNA; miRNA; aberrant gene function;
KW   gain-of-function mutation; neurological disease.
XX
OS   Synthetic.
XX
PN   WO2005007877-A2.
XX
PD   27-JAN-2005.
XX
PF   19-JUL-2004; 2004WO-US023789.
XX
PR   18-JUL-2003; 2003US-0488510P.
XX
PA   (UYMA-) UNIV MASSACHUSETTS.
XX
PI   Xu Z, Xia X;
XX
DR   WPI; 2005-112803/12.
XX
PT   New construct comprises a small hairpin
PT   RNA (shRNA) encoding nucleic acid sequence, useful for regulating shRNA
PT   expression in cells or animals.
XX
PS   Example 6; SEQ ID NO 9; 88pp; English.
XX
CC   This sequence represents the construct UbCP-lox-mirMSOD2-EGFP-lox-RFP. In
CC   this construct, the Ubiquitin C promoter (UbCP) normally directs
CC   synthesis of the hairpin and EGFP. However, upon exposure to cre, the
CC   hairpin and EGFP re excised thus preventing hairpin expression, while the
CC   RFP gene is under the control of the UbCP. This sequence is a
CC   complementary vector to UbCP-lox-RFP-lox-mirMSOD2-EGFP. The construct of
CC   the invention comprises a U6 promoter operably linked to a shRNA encoding
CC   nucleic acid sequence. The construct further comprises a first loxP site
CC   upstream of the promoter and a second loxP site downstream of the shRNA
CC   encoding sequence, the loxP sites are in the same orientation so that the
CC   promoter and encoding sequences are excisable upon exposure to Cre. The
CC   shRNA encoding sequence comprises a first stem-encoding portion, a loop-
CC   encoding portion, and a second stem-encoding portion, the loop-encoding
CC   portion comprising a first loxP site operably linked to a transcription
CC   termination signal upstream of a spacer DNA and a second loxP site, the
CC   loxP sites are in the same orientation so that the first loxP site,
CC   termination signal and spacer DNA are excisable upon exposure to Cre. The
CC   construct further comprises spacer DNA downstream of the shRNA encoding
CC   sequence, a second loxP site downstream of the spacer DNA, and a first
CC   loxP site within the loop-encoding portion of the shRNA encoding
CC   sequence, the loxP sites are in the same orientation so that the spacer
CC   DNA and second stem-encoding sequence are excisable upon exposure to Cre.
CC   The U6 promoter comprises a distal sequence element (DSE), a proximal
CC   sequence element (PSE), and a TATA box, operably linked, where the
CC   construct comprises a first loxP site downstream of the shRNA encoding
CC   sequence, and a second loxP site between the PSE and the TATA box are
CC   excisable upon exposure to Cre. The construct also comprises a first loxP
CC   site and a second loxP site, the sites are interrupted by spacer DNA
CC   between the DSE and the PSE, the loxP sites are in the same orientation
CC   so that a loxP site and the spacer DNA are excisable upon exposure to
CC   Cre. The construct is useful for the treatment of a disease, which is
CC   caused by aberrant gene function. The disease is a dominant, gain-of-
CC   function mutation, e.g. a neurological disease. The constructs are useful
CC   for regulating shRNA expression in cells or animals. The constructs,
CC   methods, kits, and compositions are useful for investigating gene
CC   functions, both physiologic and pathologic, in specific cell groups and
CC   in specific ages, in normal and diseased pathways.
XX
SQ   Sequence 4200 BP; 889 A; 1094 C; 1265 G; 950 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:   1
Score: 1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 14

Length: 4200
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

8.85e-136
DB:

US-10-006-922A-12 (1-225) x ADW43589 (1-4200)

Qy   1 MetArgSerSerLysSerAsnValIleLysGluGluPheMetArgPheLysValArgMetGluGlyGly 20
Db 3267 ATGCGCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3326

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db 3327 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3386

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 3387 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3446

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 3447 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3506

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 3507 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3566

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyCysPheIleTyr 120
Db 3567 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3626

Qy 121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysGlyThr 140
Db 3627 AAGGTGAAGTTCATCGGCGTGGCGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGGGCACC 3686

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValValLeuLysGlyGlu 160
Db 3687 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGGTGCTGAAGGGCGAG 3746

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 3747 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3806

```
Qy    181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db   3807 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGGCTACTACTACGTGGACAGCAAGCTGGAC 3866

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGlnIleTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3867 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGATACGAGCGCACGAGGGCCGC 3926

Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   3927 CACCACCTGTTCCTG 3941
```

```
AC    
AC    ADW43588;
ID    ADW43588 standard; DNA; 4300 BP.
XX
AC    ADW43588;
XX
DT    07-APR-2005  (first entry)
XX
DE    UbCP-lox-RFP-lox-mirMSOD2-EGFP.
XX
PF    19-JUL-2004; 2004WO-US023789.
PR    18-JUL-2003; 2003US-0488510P.
XX
PA    (UYMA-) UNIV MASSACHUSETTS.
XX
PI    Xu Z, Xia X;
XX
DR    WPI; 2005-112883/12.
XX
PT    New construct comprises a U6 promoter operably linked to a small hairpin
PT    RNA (shRNA) encoding nucleic acid sequence, useful for regulating shRNA
PT    expression in cells or animals.
XX
PS    Example 6; SEQ ID NO 8; 88pp; English.
XX
CC    This sequence represents the construct UbCP-lox-RFP-lox-mirMSOD2-EGFP. In
CC    this construct, the Ubiquitin C promoter (UbCP) normally transcribes RFP
CC    mRNA, which terminates at a polyA site. However, because the RFP gene is
CC    flanked by two loxP sites, upon exposure to Cre, the RFP gene is excised
CC    and the construct is converted to UbCP-lox-mirMSOD2-EGFP. The resulting
CC    recombined sequence transcribes a message that contains an intron with an
CC    shRNA (small hairpin) against the mouse SOD2 gene and an EGFP gene. The
CC    construct of the invention comprises a U6 promoter operably linked to a
CC    shRNA encoding nucleic acid sequence. The construct further comprises a
CC    first loxP site upstream of the promoter and a second loxP site
CC    downstream of the shRNA encoding sequence, the loxP sites are excisable
CC    orientation so that the promoter and encoding sequences are excisable
CC    upon exposure to Cre. The shRNA encoding sequence comprises a first stem-
CC    encoding portion, a loop-encoding portion, and a second stem-encoding
CC    portion, the loop-encoding portion comprising a first loxP site operably
CC    linked to a transcription termination signal upstream of a spacer DNA and
CC    a second loxP site, termination signal and spacer DNA are excisable upon
CC    exposure to Cre. The construct further comprises spacer DNA downstream of
CC    the shRNA encoding sequence, a second loxP site downstream of the spacer
CC    DNA, and a first loxP site within the loop-encoding portion of the shRNA
CC    encoding sequence, the loxP sites are in the same orientation so that the
CC    spacer DNA and second stem-encoding sequence a distal sequence element
CC    to Cre. The U6 promoter comprises a distal sequence element (DSE), a
CC    proximal sequence element (PSE), and a TATA box, operably linked, where
CC    the construct comprises a first loxP site downstream of the shRNA
CC    encoding sequence, and a second loxP site between the DSE and the PSE,
CC    the loxP sites are in the same orientation so that the shRNA encoding
CC    sequences and a portion of the promoter comprising the PSE and the TATA
CC    box are excisable upon exposure to Cre. The construct also comprises a
CC    first loxP site and a second loxP site, the sites are interrupted by
CC    spacer DNA, between the DSE and the PSE, the loxP sites are in the same
CC    orientation so that a loxP site and the spacer DNA are excisable upon
CC    exposure to Cre. The construct is useful for the treatment of a disease,
CC    which is caused by aberrant gene function. The disease is a dominant,
CC    gain-of-function mutation, e.g. a neurological disease. The constructs
CC    are useful for regulating shRNA expression in cells or animals. The
CC    constructs, methods, kits, and compositions are useful for investigating
CC    gene functions, both physiologic and pathology, in specific cell groups
CC    and in specific aged, in normal and diseased pathways.
XX
SQ    Sequence 4300 BP; 906 A; 1128 C; 1297 G; 967 T; 0 U; 2 Other;
```

```
Qy    181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db   3807 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGGCTACTACTACGTGGACAGCAAGCTGGAC 3866
```

```
RESULT 56
```

```
Db   3867 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGATACGAGCGCACGAGGGCCGC 3926
```

```
Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGlnIleTyrGluArgThrGluGlyArg 220
```

```
OS    Synthetic.
XX
PN    WO2005007877-A2.
XX
PD    27-JAN-2005.
XX
```

```
KW    gain-of-function mutation; neurological disease.
KW    ss; hairpin; Neuroprotective; shRNA; miRNA; aberrant gene function;
```

US-10-006-922A-12 (1-225) x ADW43588 (1-4300)

Alignment Scores:
Pred. No.:
Score: 9.15e-136
Percent Similarity: 1186.00
Best Local Similarity: 98.22%
Query Match: 97.33%
                            97.69%
                            14

Length: 4300
Matches: 219
Conservative: 2
Mismatches: 4
Indel: 0
Gaps: 0

```
Qy      1 MetArgSerSerLysAsnValIleLysGluMetMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1291 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCCCATGGAGGGC 1350

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1410

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1411 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1470

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1471 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1530

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1531 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1590

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1591 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1650

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1651 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1710

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1711 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1770

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1771 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1830

Qy    181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db   1831 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGGCTACTACTACGTGGACAGCAAGCTGGAC 1890

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGlnIleTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1831 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGATACGAGCGCACGAGGGCCGC 1950

Qy    221 HisHisLeuPheLeu 225
```

```
Db        1951 CACCACCTGTTCCTG 1965
                |||||||||||||||||

RESULT 57

ID   ADQ91220       standard; DNA; 4488 BP.
XX
AC   ADQ91220;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Fusion construct of human DCX promoter and the DSRed2 gene SeqID 6.
XX
KW   human; neuronal marker; doublecortin; DCX; ds; DSRed2; chimeric; bcl-2;
KW   PTEN; dopaminergic determination factor; Nurr1; neuronal migration;
KW   neurophilin; SDF-1; NeuroD; BMP4; transgenic; neurogenic;
KW   neural stem cell activity; neuronal differentiation;
KW   neurological disorder; epilepsy; stroke; learning; memory skill;
KW   neurodegenerative disorder; Parkinson's; Huntington's; HIV-dementia;
KW   psychiatric disorder; tumourous disorder; brain malfomation; vasotropic;
KW   anticonvulsant; cerebroprotective; antiparkinsonian; neuroprotective;
KW   nootropic; anti-HIV; regulatory sequence;
KW   proliferative neuronal determined cell; anti-apoptotic.
XX
OS   Homo sapiens.
OS   Chimeric.
OS   Unidentified.
XX
PN   WO200406775I-A1.
XX
PD   12-AUG-2004.
XX
PF   28-JAN-2004; 2004WO-EP000760.
XX
PR   28-JAN-2003; 2003EP-00002027.
XX
PA   (UYRB-) UNIV KLINIKUM REGENSBURG.
XX
PI   Couillard-Despres S, Karl C, Kuhn H, Aigner L;
XX
DR   WPI; 2004-581007/56.
XX
PS   Claim 20; SEQ ID NO 6; 270pp; English.
XX
PT   Use of a regulatory sequence for the early, transient expression of a
PT   heterologous nucleotide sequence in proliferative neuronal determined
PT   cells.
XX
CC   This invention relates to the novel use of regulatory sequences for the
CC   early, transient expression of a heterologous nucleotide sequences in
CC   proliferative neuronal determined cells. Specifically, it refers to the
CC   preparation of a recombinant vector comprising a heterologous gene
CC   sequence such as an anti-apoptotic gene e.g. bcl-2 or PTEN, a
CC   determination/differentiation gene such as the dopaminergic determination
CC   factor Nurr1, a gene capable of inducing neuronal migration e.g.
CC   neurophilin or SDF-1 or a transcription factor gene such as NeuroD or
CC   BMP4. The present invention describes non-human transgenic animals that
CC   have been transfected with such a vector that can be used for screening
CC   compounds in vitro to identify those with neurogenic properties capable
CC   of regulating neural stem cell activity, neurogenesis and/ or neuronal
CC   differentiation, which can subsequently be used in the development of
CC   appropriate pharmaceutical compositions. Furthermore, these transgenic
CC   animals can be used for the in vivo tracking of newly generated neurons,
CC   as well as for transplanted or migrating neuronal determined cells.
CC   Accordingly, such compositions can be used in gene therapy for
CC   neurological disorders e.g. epilepsy or stroke and for the enhancement of
CC   learning and/or memory skills caused by neurodegenerative disorders such
CC   as Parkinson's, Huntington's or HIV-dementia, as well as psychiatric
CC   disorders, tumourous disorders of the nervous system or a developmental
CC   or brain malfomation. Hence, these compositions exhibit vasotropic,
CC   anticonvulsant, cerebroprotective, antiparkinsonian, neuroprotective,
CC   nootropic and anti-HIV activities. This polynucleotide sequence is the
CC   fusion construct DNA of the reporter gene DSRed2 under the control of the
CC   human neuronal marker doublecortin (DCX) promoter, given in an
CC   exemplification of the invention.
XX
SQ   Sequence 4488 BP; 1184 A; 1026 C; 973 G; 1305 T; 0 U; 0 Other;
```

```
US-10-006-922A-12 (1-225) x ADQ91220 (1-4488)

Alignment Scores:
Pred. No.:                        9.72e-136
Score:                            1186.00
Percent Similarity:               98.22%      Length:             4488
Best Local Similarity:            97.33%      Matches:            219
Query Match:                      97.69%      Conservative:       2
DB:                               13          Mismatches:         4
                                              Indels:             0
                                              Gaps:               0

Qy      1  MetArgSerLysSerLysAsnValIleLeuLysGluPheMetAlaArgPheLysValArgMetGluGly  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3572  ATGGCGTCTTCCAAGAGCAAGAATGTCATCACCGAGTTCATGGCTTCAAGGTCGCATGGAGGGC  3631

Qy     21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
           |||||||||||||||||||||||||||||||||||||||||||||||||||::||||||||||
Db   3632  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  3691

Qy     41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3692  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  3751

Qy     61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3752  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  3811

Qy     81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3812  GACGACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  3871

Qy    101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3872  GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  3931

Qy    121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3932  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC  3991

Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3992  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  4051

Qy    161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4052  ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  4111

Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
           |||||||||||||||||||||||||||||||||||||||||::::|||||||||||||||||
Db   4112  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCTGAAG  4171

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluIleTyrGluMetArgThrGluGlyArg  220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4172  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCTACGAGATGCGCACCGAGGGCCGC  4231

Qy    221  HisHisHisLeuPheLeu  225
           ||||||||||||||||
Db   4232  CACCACCTGTTCCTG  4246
```

```
RESULT 58

ID   ADL35208       standard; DNA; 6706 BP.
XX
AC   ADL35208;
XX
DT   03-JUN-2004  (first entry)
XX
DE   Plasmid pLL3.2 nucleotide sequence SEQ ID NO:4.
XX
```

```
KW   lentiviral vector; functional packaging signal; multiple cloning site;
KW   antimicrobial; gene therapy; infection; plasmid; pLL3.2; circular; gene;
KW   de.
XX
OS   Human immunodeficiency virus 1.
FH
OS   Synthetic.
XX
FH   Key              Location/Qualifiers
FT
FT   promoter         212..816
FT                    /*tag= a
FT                    /note= "CMV promoter/enhancer 1"
FT   promoter         835..1509
FT                    /*tag= b
FT                    /note= "5' HIV R-U5-del gag (HIV NL4-3/454-1126)"
FT   misc_feature     1539..2396
FT                    /*tag= c
FT                    /note= "HIV RRE (HIV NL4-3/7622-8459)"
FT   misc_feature     2422..2599
FT                    /*tag= d
FT                    /note= "HIV Flap"
FT   misc_feature     2710..2745
FT                    /*tag= e
FT                    /note= "HIV Flap"
FT   CDS              2772..3452
FT                    /*tag= f
FT                    /product= "dsRed2"
FT   misc_feature     3506..3539
FT                    /*tag= g
FT                    /note= "Loxp"
FT   misc_feature     3594..4183
FT                    /*tag= h
FT                    /note= "WRE element"
FT   misc_feature     4203..4892
FT                    /*tag= i
FT                    /note= "HIV Flap"
FT   CDS              4895..5568
FT                    /*tag= j
FT                    /note= "puc"
FT   rep_origin       5713..6573
FT                    /*tag= k
FT                    /product= "AmpR"
CDS                   /note= "3' SIN LTR"
```

WO2004022722-A2.

PN   18-MAR-2004.

PD   05-SEP-2003; 2003WO-US028111.

PF   06-SEP-2002; 2002US-0408558P.
PR   27-SEP-2002; 2002US-0414195P.
PR   21-NOV-2002; 2002US-0428039P.

PR   Claim 43; SEQ ID NO 4; 197pp; English.

PS   (MASI) MASSACHUSETTS INST TECHNOLOGY.
PA

PI   Bear JE, Dillon CP, Rubinson DA, Van Parijs L;

XX   WPI; 2004-315566/29.
DR
XX
XX   New lentiviral vector comprising a nucleic acid whose sequence includes a
PT   functional packaging signal, a multiple cloning site (MCS), and at least
PT   one additional element, useful for treating infectious diseases.

The present invention describes a lentiviral vector comprising a nucleic
acid whose sequence includes a functional packaging signal, a multiple
cloning site (MCS), and at least one additional element selected from a
second MCS, a heterologous nucleic acid is
inserted, an HIV FLAP element, an expression-enhancing
posttranscriptional regulatory element, a target site for a site-specific
recombinase, and a self-inactivating (SIN) LTR. Also described: (1) a
lentiviral vector comprising the RNA polymerase III promoter; (2) a
collection of at least two lentiviral vectors; (3) a three plasmid

```
CC   lentiviral expression system; (4) a four plasmid lentiviral expression
CC   system comprising the three plasmid lentiviral expression system, further
CC   comprising a four plasmid comprising a nucleic acid sequence that encodes
CC   Rev, operably linked to a promoter; (5) a cell comprising the lentiviral
CC   vector; (6) a cell comprising a provirus derived from the lentiviral
CC   vector; (7) a transgenic animal, at least some of whose cells contain a
CC   provirus derived from the lentiviral vector; (8) a method of creating a
CC   producer cell line; (9) a method of producing lentiviral particles; (10)
CC   a method of expressing a heterologous nucleic acid in a target cell; (11)
CC   a method for achieving controlled expression of a heterologous nucleic
CC   acid in a cell; (12) a method for expressing a transcript in a mammal in
CC   a cell type or tissue-specific manner; (13) a composition comprising the
CC   lentiviral vector, and a delivery agent that enhances delivery of the
CC   vector to cells; (14) a pharmaceutical composition comprising the
CC   lentiviral vector and a carrier; (15) methods of reversibly inhibiting or
CC   reducing expression of a target transcript in a cell, or in a mammal in a
CC   cell type or tissue-specific manner; (16) a method of treating or
CC   preventing infection by an infectious agent, or a disease or clinical
CC   condition; and (17) a kit comprising the lentiviral transfer plasmid and
CC   a packaging mix. The lentiviral vector has antimicrobial activity, and
CC   can be used in gene therapy. The lentiviral vector, compositions and
CC   methods are useful for treating infections. The present sequence
CC   represents the plasmid pLL3.2 nucleotide sequence, which is used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 6706 BP; 1755 A; 1638 C; 1722 G; 1591 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:
Score: 1.71e-135   Length: 6706
Percent Similarity: 1186.00   Matches: 219
Best Local Similarity: 98.22%   Conservative: 2
Query Match: 97.33%   Mismatches: 4
DB: 97.69%   Indels: 0
                    12   Gaps: 0

US-10-006-922A-12 (1-225) x ADL35208 (1-6706)

```
Qy     1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2778 ATGGCCTCCTTCCGAGAAGCGTCATCACCGAGTTCATGCGCTTCAAGGTGCCATGGAGGGC 2837

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2838 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2897

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2898 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2957

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2958 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3017

Qy    81 AspThrTyrLysLeuLysSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3018 GACTACAAGAGAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3077

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3078 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3137

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3138 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 3197

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3198 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3257

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3258 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3317
```

```
Qy    181  TyrMetAlaLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3318  TACATGGCCAAGAGACCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 3377

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3378  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 3437

Qy    221  HisHisLeuPheLeu 225
           |||||||||||||||
Db   3438  CACCACCTGTTCCTG 3452
```

RESULT 59

```
ID   ADR30815  standard; DNA; 7495 BP.
XX
AC   ADR30815;
XX
DT   04-NOV-2004 (first entry)
XX
DE   Zebrafish BDNF gene construct DNA #3.
XX
KW   Brain-derived neurotrophic factor; BDNF; promoter; zebrafish; ds.
XX
OS   Danio rerio.
OS   Chimeric.
OS   Unidentified.
XX
FH   Key            Location/Qualifiers
FH
FT   misc_feature   1..20
FT                  /*tag= a
FT                  /note= "Vector DNA"
FT   misc_feature   15..5091
FT                  /*tag= b
FT                  /note= "Expression vector DNA"
FT   misc_feature   21..1776
FT                  /*tag= c
FT                  /note= "Zebrafish BDNF gene 5' flank region"
FT   promoter       1777..1804
FT                  /*tag= d
FT                  /note= "zebrafish BDNF promoter gene"
FT   exon           1805..2099
FT                  /*tag= e
FT                  /note= "Exon 1c (5' UT) of Zebrafish BDNF gene "
FT   misc_feature   2100..2122
FT                  /*tag= f
FT                  /note= "Vector DNA"
FT   misc_feature   2123..2800
FT                  /*tag= g
FT                  /note= "Reporter vector DNA"
FT   misc_feature   2801..2807
FT                  /*tag= h
FT                  /note= "Linker DNA"
FT   misc_feature   2808..5086
FT                  /*tag= i
FT                  /note= "3' flank region"
FT   misc_feature   5087..7495
FT                  /*tag= j
FT                  /note= "Vector DNA"
XX
```

```
DR   WPI; 2004-592772/57.
XX
PT   Nucleic acid construct useful for screening substances for their ability
PT   to influence production of brain derived neurotrophic factor (BDNF),
PT   comprises promoter of BDNF gene and red fluorescent reporter.
XX
PS   Claim 26; SEQ ID NO 5; 56pp; English.
XX
CC   The present invention provides an isolated brain-derived neurotrophic
CC   factor (BDNF) gene promoter and its associated nucleic acid constructs.
CC   The invention is useful for screening a number of substances for their
CC   ability to influence the production of BDNF in living organisms. The
CC   present sequence is zebrafish brain-derived neurotrophic factor (BDNF)
CC   gene construct DNA. This sequence is used in the invention.
XX
SQ   Sequence 7495 BP; 2249 A; 1599 C; 1659 G; 1988 T; 0 U; 0 Other;
```

US-10-006-922A-12 (1-225) x ADR30815 (1-7495)

Alignment Scores:

Pred. No.:	.	2e-135	Length: 7495
Score:	1186.00	Matches:	219
Percent Similarity:	98.22%	Conservative:	2
Best Local Similarity:	97.33%	Mismatches:	4
Query Match:	97.69%	Indels:	0
DB:	13	Gaps:	0

```
Qy      1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           |||  ::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2123  ATGGCTCGTCCTCCGAGAACGGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGGC 2182

Qy     21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2183  ACCGTGAACGGCCACGAGTTCGAGATCGAAGGGGGAGGGGAGGGCCGCCCCTACGAGGGC 2242

Qy     41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2243  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2302

Qy     61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2303  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCAGATCCCC   2362

Qy     81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2363  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2422

Qy    101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||| :||||||||||||||
Db   2423  GACGGCGGCGTGGTGACCGTGACCCAGGACTCTCCCTGCAGGACGGCTGCTTCATCTAC  2482

Qy    121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2483  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 2542

Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2543  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2602

Qy    161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2603  ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGTCTATC 2662

Qy    181  TyrMetAlaLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
           |||||||||| :|||||||||||||||||||||||||||||||||||||||||||||
Db   2663  TACATGGCCAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC   2722

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2723  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2782

Qy    221  HisHisLeuPheLeu 225
           |||||||||||||||
```

```
Db    2783 CACCACCTGTTCCTG 2797
```

RESULT 60
XX
ID   ADQ91222 standard; DNA; 7616 BP.
XX
AC   ADQ91222;
XX
DT   04-NOV-2004 (first entry)
XX
DE   Expression vector phuDCXpromoDSRed2 DNA SeqID 8.
XX
KW   human; neuronal marker; doublecortin; DCX; da; DSRed2; expression vector;
KW   bcl-2; PTEN; dopaminergic determination factor; Nurr1;
KW   neuronal migration; neurophilin; SDF-1; NeuroD; BMP4; transgenic;
KW   neurogenic; neural stem cell activity; neuronal differentiation;
KW   neurological disorder; epilepsy; stroke; learning; memory skill;
KW   neurodegenerative disorder; tumourous disorder; Parkinson's; HIV-dementia;
KW   psychiatric disorder; cerebroprotective; brain malformation; vasotropic;
KW   anticonvulsant; cerebroprotective; antiparkinsonian; neuroprotective;
KW   nootropic; anti-HIV; regulatory sequence;
KW   proliferative neuronal determined cell; anti-apoptotic.
XX
OS   Homo sapiens.
OS   Synthetic.
OS   Unidentified.
XX
PN   WO200467751-A1.
XX
PD   12-AUG-2004.
XX
PF   28-JAN-2003; 2003EP-00002027.
XX
PR   28-JAN-2004; 2004WO-EP000760.
XX
PA   (UYRB-) UNIV KLINIKUM REGENSBURG.
XX
PI   Couillard-Despres S, Karl C, Kuhn H, Aigner L;
XX
DR   WPI; 2004-581007/56.
XX
PT   Use of a regulatory sequence for the early, transient expression of a
PT   heterologous nucleotide sequence in proliferative neuronal determined
PT   cells.
XX
PS   Claim 20; SEQ ID NO 8; 270pp; English.
XX
CC   This invention relates to the novel use of regulatory sequences for the
CC   early, transient expression of a heterologous nucleotide sequences in
CC   proliferative neuronal determined cells. Specifically, it refers to the
CC   preparation of a recombinant vector comprising a heterologous gene
CC   sequence such as an anti-apoptotic gene e.g. bcl-2 or PTEN, a
CC   determination/differentiation gene such as the dopaminergic determination
CC   factor Nurr1, a gene capable of inducing neuronal migration e.g.
CC   neurophilin or SDF-1 or a transcription factor gene such as Neurod or
CC   BMP4. The present invention describes non-human transgenic animals that
CC   have been transfected with such a vector that can be used for screening
CC   compounds in vitro to identify those with neurogenic properties capable
CC   of regulating neural stem cell activity, neurogenesis and/ or neuronal
CC   differentiation, which can subsequently be used in the development of
CC   appropriate pharmaceutical compositions. Furthermore, these transgenic
CC   animals can be used for the in vivo tracking of newly generated neurons,
CC   as well as for transplanted or migrating neuronal determined cells.
CC   Accordingly, such compositions can be used in gene therapy for
CC   neurological disorders e.g. epilepsy or stroke and for the enhancement of
CC   learning and/or memory skills caused by neurodegenerative disorders such
CC   as Parkinson's, Huntington's or HIV-dementia, as well as psychiatric
CC   disorder, tumourous disorders of the nervous system or a developmental
CC   or brain malformation. Hence, these compositions exhibit vasotropic,
CC   anticonvulsant, cerebroprotective, antiparkinsonian, neuroprotective,
CC   nootropic and anti-HIV activities. This polynucleotide sequence is the
CC   expression vector phuDCXpromoDSRed2, which comprises the reporter gene
CC   DSRed2 under the control of the human neuronal marker doublecortin (DCX)
CC   promoter, given in an exemplification of the invention.
XX
SQ   Sequence 7616 BP; 1916 A; 1854 C; 1822 G; 2024 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADQ91222 (1-7616)
DB:

Alignment Scores:
Pred. No.:   2.05e-135   Length: 7616
Score: 1186.00   Matches: 219
Percent Similarity: 98.22%   Conservative: 2
Best Local Similarity: 97.33%   Mismatches: 4
Query Match: 97.69%   Indels: 0
DB: 13   Gaps: 0
```

```
Qy       1 MetArgSerSerLysAsnValIleLysGlyGluPheMetArgPheLysValArgMetGluGly 20
             |||::::|||::||||||||||||||||||||||||||||||||||||||||||||
Db    3606 ATGCGCTCTTCCAAGAACGTCATCAAGGGCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3665

Qy      21 ThrValAsnGlyHisGlyPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3666 ACCGTGAACGGCCACGGATTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3725

Qy      41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3726 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3785

Qy      61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3786 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3845

Qy      81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3846 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 3905

Qy     101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||::|||||
Db    3906 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3965

Qy     121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3966 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 4025

Qy     141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4026 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 4085

Qy     161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4086 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 4145

Qy     181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           ||||||||||||||||||||||||||::|||||||||||||||||||||||||||||||
Db    4146 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 4205

Qy     201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4206 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 4265

Qy     221 HisHisLeuPheLeu 225
           |||||||||||||||
Db    4266 CACCACCTGTTCCTG 4280
```

RESULT 61
XX
ID   ADL35211 standard; DNA; 7927 BP.
XX
AC   ADL35211;
XX
DT   03-JUN-2004 (first entry)
XX
DE   Plasmid pLL3.5 nucleotide sequence SEQ ID NO:7.
XX
KW   lentiviral vector; functional packaging signal; multiple cloning site;

KW   antimicrobial; gene therapy; infection; plasmid; pLL3.5; circular; gene;
KW   ds.
OS   Human immunodeficiency virus 1.
OS   Synthetic.

FH   Key   Location/Qualifiers
FT   promoter   212..816
FT   /tag= a
FT   /note= "CMV promoter/enhancer 1"
FT   misc_feature   835..1509
FT   /tag= b
FT   /note= "5' HIV R-U5-del gag (HIV NL4-3/7622-8459)"
FT   misc_feature   1539..2396
FT   /tag= c
FT   /note= "HIV Flap"
FT   promoter   2422..2599
FT   /tag= d
FT   /note= "HIV RRE (HIV NL4-3/454-1126)"
FT   misc_feature   2632..3841
FT   /tag= e
FT   /note= "UbC promoter"
FT   promoter   3891..3966
FT   /tag= f
FT   /note= "HIV Flap"
FT   misc_feature   3993..4673
FT   /tag= g
FT   /product= "dsRed2"
FT   CDS
FT   /note= "3'"
FT   /tag= j
FT   /note= "SIN LTR"
FT   rep_origin   6116..6789
FT   /tag= k
FT   /note= "pUC"
FT   misc_feature   4727..4760
FT   /tag= h
FT   /note= "Loxp"
FT   misc_feature   4815..5404
FT   /tag= i
FT   /note= "WRE element"
FT   misc_feature   5424..6113
FT   /tag=
FT   /note= "3"
FT   CDS   6934..7794
FT   /tag= 1
FT   /product= "AmpR"

PS

The present invention describes a lentiviral vector comprising a nucleic
acid whose sequence includes a functional packaging signal, a multiple
cloning site (MCS), and at least one additional element selected from a
second MCS, a second MCS into which a heterologous nucleic acid is
inserted, an HIV FLAP element, an expression-enhancing
posttranscriptional regulatory element, a target site for a site-specific
recombinase, and a self-inactivating (SIN) LTR. Also described: (1) a

CC   lentiviral vector comprising the RNA polymerase III promoter; (2) a
CC   collection of at least two lentiviral vectors; (3) a three plasmid
CC   lentiviral expression system; (4) a four plasmid lentiviral expression
CC   system comprising the three plasmid lentiviral expression system, further
CC   comprising a four plasmid comprising a nucleic acid sequence that encodes
CC   Rev, operably linked to a promoter; (5) a cell comprising the lentiviral
CC   vector; (6) a cell comprising a provirus derived from the lentiviral
CC   vector; (7) a transgenic animal, at least some of whose cells contain a
CC   provirus derived from the lentiviral vector; (8) a method of creating a
CC   producer cell line; (9) a method of producing lentiviral particles; (10)
CC   a method of expressing a heterologous nucleic acid in a target cell; (11)
CC   a method for achieving controlled expression of a heterologous nucleic
CC   acid in a cell; (12) a method for expressing a transcript in a mammal in
CC   vector; (14) a method for expressing a transcript in a mammal in
CC   a cell type or tissue-specific manner; (13) a composition comprising the
CC   lentiviral vector, and a delivery agent that enhances delivery of the
CC   vector to cells; (14) a pharmaceutical composition comprising the
CC   lentiviral vector and a carrier; (15) methods of reversibly inhibiting or
CC   reducing expression of a target transcript in a cell, or in a mammal in a
CC   cell type or tissue-specific manner; (16) a method of treating or
CC   preventing infection by an infectious agent, or a disease or clinical
CC   condition; and (17) a kit comprising the lentiviral transfer plasmid and
CC   a packaging mix. The lentiviral vector has antimicrobial activity, and
CC   methods are useful for treating infections. The present sequence
CC   represents the plasmid pLL3.5 nucleotide sequence, which is used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 7927 BP; 1958 A; 1925 C; 2151 G; 1893 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x ADL35211 (1-7927)

Alignment Scores:
Pred. No.:   Length: 7927
Score: 1186.00   Matches: 219
Percent Similarity: 98.22%   Conservative: 2
Best Local Similarity: 97.33%   Mismatches: 4
Query Match: 12   Indels: 0
                                    Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGlyGly 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3999 ATGGCCTCCTCCGAGAACGTCATCACCCAGTTCATGCCCTTCAAGGTTCGCATGGAGGGC 4058

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4059 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCATACGAGGGC 4118

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4119 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 4178

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4179 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 4238

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4239 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 4298

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4299 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 4358

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4359 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 4418

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuGlyGlu 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4419 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGCTGGGCGAG 4478

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            ||||||||||||||||||||||||||||||

```
Db    4479 ACCCACAAGGCCCTGGAAGCTGAAGGACGGCGGCCCACTACCTGGTGGGAGTTCAAGTCCATC 4538

Qy     181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLysAsp 200
           |||||||||||||||||||||||||||||||||||||||::::|||||||||||||||
Db    4539 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACCGCCAAGCTGGAC 4598

Qy     201 IleThrSerHisAsnGlyAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    4599 ATCACCTCCCACAACGGAGACTACACCATCGTGCAGTACGAGCGCACCGAGGGCCGC 4658

Qy     221 HisHisLeuPheLeu 225
           |||||||||||||||
Db    4659 CACCACCTGTTCCTG 4673
```

```
RESULT 62
ID      ADM82791        standard; cDNA; 9731 BP.
XX
AC      ADM82791;
XX
DT      03-JUN-2004  (first entry)
XX
DE      DNA repair pathway related retroviral vector cDNA with CMV promoter.
XX
KW      inducer; inhibitor; DNA repair pathway; anti-HIV; cytostatic; virucide;
KW      antidiabetic; neuroprotective; retroviral infection; AIDS; HIV infection;
KW      cancer; human adult T-cell leukaemia; lymphoma;
KW      feline immunodeficiency virus; Type I diabetes; multiple sclerosis;
KW      gene therapy; human; cyclic; circular; CMV promoter; 88.
XX
OS      Unidentified.
XX
PN      WO200309573-A2.
XX
PD      30-OCT-2003.
XX
PF      04-APR-2003; 2003WO-US010302.
XX
PR      05-APR-2002; 2002US-0370376P.
XX
PI      Fishel RA, Yoder KE;
XX
DR      WPI; 2003-854096/79.
XX
PA      (FISH/) FISHEL R A,
PA      (YODE/) YODER K E.
XX
PT      Screening for compounds that modulate a DNA repair pathway and/or
PT      retroviral integration, useful for treating retroviral infection,
PT      comprises determining the amount of a retroviral cDNA circularization in
PT      the presence of the test compound.
XX
PS      Claim 73; SEQ ID NO 5; 89pp; English.
XX
```

The invention relates to a novel method for screening for inducers or
inhibitors of a DNA repair pathway by contacting at least one component
of a DNA repair pathway with a non-circularized retroviral cDNA in the
presence and absence of a test compound, and determining whether
circularization of the cDNA is increased or decreased in the presence of
the test compound. The DNA repair pathway components have the following
activities: anti-HIV, cytostatic, virucide, antidiabetic, and
neuroprotective. The method is useful for identifying compounds that
modulate a DNA repair pathway and/or retroviral activity. The compound is
used in manufacturing a pharmaceutical composition for the treatment of a
of a DNA repair pathway infection (e.g. AIDS, HIV infection, cancer, human adult T-
cell leukaemia, lymphoma, feline immunodeficiency virus, Type I diabetes
or multiple sclerosis) or for increasing the efficiency of gene delivery
in a gene therapy. This polynucleotide represents a retroviral cDNA
sequence of the invention.

```
SQ      Sequence 9731 BP; 2444 A; 2412 C; 2548 G; 2327 T; 0 U; 0 Other;
```

.Alignment Scores:

```
                                        Pred. No.;
DB:

Score:                          2.89e-135      Length:            9731
Percent Similarity:             1186.00        Matches:            219
Best Local Similarity:          98.22%         Conservative:         2
                                97.33%         Mismatches:           4
Query Match:                    97.69%         Indels:               0
                                11             Gaps:                 0
```

```
US-10-006-922a-12 (1-225) x ADM82791 (1-9731)
```

```
Qy       1 MetArgSerSerLysAsnValIleGluLysGluPheMetArgPheLysValArgMetGluGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5496 ATGGCCTTCCCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 5555

Qy      21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5556 ACCGTGGCCAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 5615

Qy      41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5616 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5675

Qy      61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5676 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5735

Qy      81 AspThrTyrLysLeuAsnProAspGlyTyrProGluGlyTrpGluMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    5736 GACTACAAGCTGAACCCCGACGGCTACCCCGAGGGCTGGGAGATGAACTTCGAG 5795

Qy     101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5796 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 5855

Qy     121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5856 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 5915

Qy     141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5916 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 5975

Qy     161 IleHisGlnAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5976 ATCCACCAAGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 6035

Qy     181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLysAsp 200
           |||||||||||||||||||||||||||||||||::::|||||||||||||||
Db    6036 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 6095

Qy     201 IleThrSerHisAsnGlyAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    6096 ATCACCTCCCACAACGGAGACTACACCATCGTGCAGTACGAGCGCACCGAGGGCCGC 6155

Qy     221 HisHisLeuPheLeu 225
           |||||||||||||||
Db    6156 CACCACCTGTTCCTG 6170
```

```
RESULT 63
ID      ADM82792        standard; cDNA; 9782 BP.
XX
AC      ADM82792;
XX
DT      03-JUN-2004  (first entry)
XX
DE      DNA repair pathway related retroviral vector cDNA with MSH2 promoter.
XX
KW      inducer; inhibitor; DNA repair pathway; anti-HIV; cytostatic; virucide;
KW      antidiabetic; neuroprotective; retroviral infection; AIDS; HIV infection;
KW      cancer; human adult T-cell leukaemia; lymphoma;
KW      feline immunodeficiency virus; Type I diabetes; multiple sclerosis;
KW      gene therapy; human; cyclic; circular; MSH2 promoter; 88.
XX
```

OS   Unidentified.
XX
PN   WO2003089573-A2.
XX
PD   30-OCT-2003.
XX
PF   04-APR-2003; 2003WO-US010302.
XX
PR   05-APR-2002; 2002US-0370376P.
XX
PA   (FISH/) FISHEL R A.
PA   (YODE/) YODER K E.
XX
PI   Fishel RA, Yoder KE;
XX
DR   WPI; 2003-854096/79.
XX
PS   Claim 73; SEQ ID NO 6; 89pp; English.

CC
CC   The invention relates to a novel method for screening for inducers or
CC   inhibitors of a DNA repair pathway by contacting a retroviral cDNA in the
CC   presence and absence of a test compound, and determining whether
CC   circularization of the cDNA is increased or decreased in the presence of
CC   the test compound. The DNA repair pathway components have the following
CC   activities: anti-HIV, cytostatic, virucide, antidiabetic, and
CC   neuroprotective. The method is useful for identifying compounds that
CC   modulate a DNA repair pathway and/or retroviral activity. The compound is
CC   used in manufacturing a pharmaceutical composition for the treatment of a
CC   retroviral infection (e.g. AIDS, HIV infection, cancer, human adult T-
CC   cell leukaemia, lymphoma, feline immunodeficiency virus, Type I diabetes
CC   or multiple sclerosis) or for increasing the efficiency of gene delivery
CC   in a gene therapy. This polynucleotide represents a retroviral cDNA
CC   sequence of the invention.
XX
SQ   Sequence 9782 BP; 2628 A; 2305 C; 2351 G; 2498 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.9le-135
Percent Similarity: 1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 97.69%
DB: 11

Length: 9782
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x ADM82792 (1-9782)

```
Qy       1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db       1 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 20

Qy      21 ThrValAlaSnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db      73 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 132

Qy      41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db     133 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 192

Qy      61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db     193 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 252

Qy      81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db     253 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 312

Qy     101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db       ...
```

Db   313 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 372

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   373 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 432

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   433 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 492

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   493 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 552

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   553 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 612

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   613 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 672

Qy   221 HisHisLeuPheLeu 225
Db   673 CACCACCTGTTCCTG 687

RESULT 64
AAA48742

ID   AAA48742   standard; cDNA; 898 BP.
XX
AC   AAA48742;
XX
DT   15-SEP-2003   (revised)
DT   19-SEP-2000   (first entry)
XX
DE   Discosoma sp. "red" novel fluorescent protein drFP583 cDNA.
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling; 88.
XX
OS   Discosoma sp; "red".
XX
FH   Key   Location/Qualifiers
FT   CDS   93..770
FT   /product= "drFP583"
FT   /transl_except= (pos:456..457,aa:Val)
FT   /transl_except= (pos:470..473,aa:Val)
FT   /note= "The codon at position 456..457 has an apparent 1
FT   nucleotide deletion which alters the reading frame. The
FT   codon at position 470..473 has an apparent 1 nucleotide
FT   insertion which corrects the altered reading frame"
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029473.
XX
PR   11-DEC-1998; 98US-00210330.
PR   14-OCT-1999; 99US-00418529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
DR   WPI; 2000-423381/36.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 6; Page 73-74; 86pp; English.

XX
CC The present sequence is the cDNA sequence of drFP583 from Discosoma sp.
CC "red", a non-bioluminescent species of the Class Anthozoa. drFP583 is a
CC full-length cDNA encoding a novel fluorescent protein (nFP). Fluorescent
CC protein can be used in fluorescent labeling, a useful tool for marking a
CC protein, cell or organism of interest. Unlike other markers used in
CC protein labeling, such as beta-galactosidase and luciferase, fluorescent
CC proteins do not require an exogenous cofactor or substrate. Methods
CC involving fluorescent proteins are also less laborious and less difficult
CC to control than the traditional methods of fluorescent labeling, where a
CC protein of interest is purified and then covalently conjugated to a
CC fluorophore derivative. Novel fluorescent proteins isolated from species
CC of the Class Anthozoa can be used as markers for gene expression and
CC protein localization studies, and in fluorescence resonance energy
CC transfer (FRET) reactions. They may have improved properties and better
CC suitability for larger excitations compared to prior art fluorescent
CC proteins such as green fluorescent protein. (Updated on 15-SEP-2003 to
CC standardise OS field)
XX
SQ Sequence 898 BP; 273 A; 172 C; 240 G; 213 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 1.78e-136
Score: 1184.00
Percent Similarity: 98.22%
Percent Similarity: 97.33%
Best Local Similarity: 97.53%
Query Match: 3

Length: 898
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

DB:

US-10-006-922A-12 (1-225) x AAA48742 (1-898)

```
Qy   1 MetArgSerSerLysGlyGluAsnValIleLeuLysGlyGluPheMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93 ATGAGGTCTTCCAAGGAAAATGTTATCAAGGAGGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 152

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly       40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153 ACGGTCAATGGGCACGAGTTTGAAATAGAGGGCGAAGGAGAGGGCCGTACCTACGAAGGC     212

Qy  41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle    60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 213 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT     272

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro       80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 273 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA     332

Qy  81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu      100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 333 GACTATATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA   392

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr      120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 393 GACGGTGGCGTCGTTACTGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC     452

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr      140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 453 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA     512

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu      160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 513 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG     572

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle      180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 573 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTGGTTGAATTCAAAAGTATT     632

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp      200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 633 TACATGGCAAAGAAGCCTGTTCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT     692

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg      220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 693 ATAACAAGCCACAACGAAGAGACTATACAATCGTTGAGCAGTTATGAAAAGACCGAGGGACGC 752

Qy 221 HisHisLeuPheLeu 225
         |||||||||||||
Db 753 CACCACCTGTTCCTT 767
```

RESULT 65
AAD03615
ID AAD03615 standard; cDNA; 898 BP.
XX
AC AAD03615;
XX
DT 11-SEP-2003 (revised)
DT 19-JUN-2001 (first entry)
XX
DE Discosoma sp. red chromo/fluorescent protein, drFP583 (NFP-6) cDNA.
XX
KW Anthozoa; Chromoprotein; fluorescent protein; drFP583; NFP-6; sunscreen;
KW analyte detection assay; selectable marker; recombinant DNA application;
KW biosensor; pH indicator; invivo marker; selective filter; ss.
XX
OS Discosoma sp; red.
XX

FH Key   Location/Qualifiers
FT CDS   93..770
FT   /tag= a
FT   /product= "Chromo/fluorescent protein, drFP583 (NFP-6)"
FT   /transl_except= (pos:540..600, aa:Leu-Gly)
FT   /note= "Insertion of 60 bases alters the reading frame"

XX
PN WO200127150-A2.
XX
PD 19-APR-2001.
XX
PF 13-OCT-2000; 2000WO-US028477.
XX
PR 14-OCT-1999; 99US-00418529.
PR 15-OCT-1999; 99US-00418917.
PR 15-OCT-1999; 99US-00418922.
PR 19-NOV-1999; 99US-00444338.
PR 19-NOV-1999; 99US-00444341.
PR 09-DEC-1999; 99US-00457556.
PR 09-DEC-1999; 99US-00457898.
PR 09-DEC-1999; 99US-00458144.
PR 09-DEC-1999; 99US-00458477.
PR 10-DEC-1999; 99WO-US229405.
PR 14-JUN-2000; 2000US-0211607P.
PR 14-JUN-2000; 2000US-0211609P.
PR 14-JUN-2000; 2000US-0211626P.
PR 14-JUN-2000; 2000US-0211627P.
PR 14-JUN-2000; 2000US-0211687P.
PR 14-JUN-2000; 2000US-0211766P.
PR 14-JUN-2000; 2000US-0211888P.
PR 14-JUN-2000; 2000US-0212070P.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Terskikh A;
XX
DR WPI; 2001-266409/27.
DR P-PSDB; AAE00375.
XX
PT An Anthozoa chromo- or fluorescent protein (P1) present in an environment
PT other than its natural environment, useful as a label in analyte
PT detection assays and as a selectable marker in recombinant DNA
PT applications.
XX
PS Claim 13; Fig 6; 69pp; English.
XX
CC The present sequence is a Discosoma sp. red chromo/fluorescent protein,
CC drFP583 (NFP-6) cDNA. NFP-6 is present in an environment other than its
CC natural environment and has an absorbance maximum ranging from 250nm to

```
CC    750nm and more usually from 540nm to 580nm and emission maximum ranging
CC    from 275nm to 775nm and more usually from 565 to 605nm. The
CC    chromoproteins or fluorescent proteins are useful as labels in analyte
CC    detection assays, as selectable markers in recombinant DNA applications,
CC    as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
CC    and as in vivo markers in animals. They are also useful in sunscreens and
CC    as selective filters. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ    Sequence 898 BP; 273 A; 173 C; 240 G; 212 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:
Score:                    1.78e-136        Length:           898
Percent Similarity:       1184.00          Matches:          219
Best Local Similarity:    98.22%           Conservative:     2
Query Match:              97.33%           Mismatches:       4
DB:                       97.53%           Indels:           0
                          4                Gaps:             0
```

US-10-006-922A-12 (1-225) x AAD03615 (1-898)

```
Qy    1    MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1    ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGTGGAGAGGA 152

Qy    21   ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db    93   ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGGAGGC 212

Qy    41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db    153  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGGCTTGGGATATT 272

Qy    61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db    213  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 332

Qy    81   AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db    273  GACTATATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTCATGAACTTTGAA 392

Qy    101  AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db    333  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 452

Qy    121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db    393  AAGGTCAAGTTCATTGGCGTTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 512

Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db    453  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 572

Qy    161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db    513  ATCCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 632

Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db    573  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGATACTATTATGTTGACTCCAAACTGGAT 692

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluLnTyrGluArgThrGluGlyArg 220
Db    633  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGCC 752

Qy    221  HisHisHisLeuPheLeu 225
Db    693  CACCATCTGTTCCTT 767
```

```
CC    The invention describes a polynucleotide sequence (I) encoding a
CC    Discosoma red fluorescent protein (DsRed) variant having a reduced
CC    propensity to oligomerize, comprising amino acid substitutions at the AB
CC    and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC    amino acids, given in the specification, where the substitutions result
CC    in reduced propensity of the DsRed variant to form tetramers. (I) is
CC    useful for detecting transcriptional activity by providing a host cells
CC    conatining a vector which comprises (I) operatively linked to an
CC    expression control sequence, and an unit to assay the variant fluorescent
CC    protein fluorescence, and assaying fluorescence of the variant
CC    fluorescent protein produced by (VII), where variant fluorescent protein
CC    fluorescence is indicative of transcriptional activity. A polynucleotide
CC    encoding a fusion protein is useful for the analysis of in vivo
CC    localisation or trafficking of a polypeptide of interest. A polypeptide
CC    marker is useful as markers to identify the location and amount of a
CC    target protein produced, where the target protein is fused to the marker,
CC    as a complement to or alternative for the green fluorescent protein or
CC    its spectral variant, for detecting induction of transcription, in
CC    applications involving fluorescence energy resonance transfer (FRET),
CC    which detects events as the function of the movement of fluorescent
CC    donors and acceptors towards or away from each other, for making
CC    fluorescent sensors for protein kinase and phosphatase activities or
CC    indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC    presence of a molecule in a sample, for identifying a specific
CC    interaction of a first and second molecule, for determining whether a
CC    sample contains an enzyme or for determining the pH of the sample. (I) is
CC    useful for identifying a region or condition that regulates the activity
CC    of an expression control sequence. This sequence encodes Discosoma red
CC    fluorescent protein variant fast T1.
XX
SQ    Sequence 678 BP; 225 C; 225 C; 205 G; 107 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:
Score:                    1.13e-133        Length:           678
Percent Similarity:       1160.00          Matches:          216
Best Local Similarity:    97.33%           Conservative:     3
                          96.00%           Mismatches:       6
```

RESULT 66

Query Match: 95.55%
DB: 10
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x ADC24128 (1-678)

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
        |||   |::|||                                                   
Db    1 ATGGCCTCCTCCGGAGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        :::||||||||                                                    
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        |||                                                           
Db  121 ACCCAGACACGGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        ||||||||||||||||||||||||||||                                  
Db  181 CTGTCCCCCCAGTTCCAGTACGGGTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81 AspTyrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||                                                           
Db  241 GACTACAACAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTCCTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy  221 HisHisHisLeuPheLeu 225
        ||||||||||||||||||
Db  661 CACCACCTGTTCCTG 675
```

RESULT 67
ADL46207

```
ID   ADL46207 standard; DNA; 678 BP.
XX
AC   ADL46207;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Discosoma red fluorescent protein (DsRed) variant T1 coding sequence.
XX
KW   red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..678
FT                  /tag= a
FT                  /product= "DsRed variant T1 protein"
XX
```

```
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PS   Disclosure; SEQ ID NO 5; 166pp; English.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   coding sequence for the variant T1.
XX
SQ   Sequence 678 BP; 141 A; 225 C; 205 G; 107 T; 0 U; 0 Other;
XX
```

Alignment Scores:

Pred. No.:	1.13e-133
Score:	1160.00
Percent Similarity:	97.33%
Best Local Similarity:	96.00%
Query Match:	95.55%

Length:	678
Matches:	216
Conservative:	3
Mismatches:	6
Indels:	0
Gaps:	0

DB: 11

US-10-006-922A-12 (1-225) x ADL46207 (1-678)

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
        |||   |::|||                                                   
Db    1 ATGGCCTCCTCCGGAGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        :::||||||||                                                    
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        |||                                                           
Db  121 ACCCAGACACGGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        ||||||||||||||||||||||||||||                                  
Db  181 CTGTCCCCCCAGTTCCAGTACGGGTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81 AspTyrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||                                                           
Db  241 GACTACAACAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTCCTTCATCTAC 360
```

```
Qy  121 LysSerValLysPheIleGlyValAsnPheProSerArgGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCGTCCCGAGGCCCCGTAATGCAGAAGAAGACT 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLysAspGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATCCACAAGGCCCTGAAGAAGGACGGGCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  661 CACCACCTGTTCCTG 675
```

RESULT 68
```
ID   ADN33980 standard; DNA; 704 BP.
XX
AC   ADN33980;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Mutant-type DsRED encoding sequence.
XX
KW   Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED;
KW   mutant; mutein; ds.
XX
OS   Discosoma sp.
XX
XX
PN   WO200305454158-A2.
XX
PD   03-JUL-2003.
XX
PF   18-DEC-2002; 2002WO-US040539.
XX
PR   19-DEC-2001; 2001US-0341723P.
XX
PA   (UYCH-) UNIV CHICAGO.
PI   Bevis B, Glick B;
XX
DR   WPI; 2003-569236/53.
XX
PT   Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT   mutant of a Cnidarian chromo- or fluorescent protein or its mutant,
PT   useful for applications involving chromo- or fluorescent proteins.
PS   Claim 7; SEQ ID NO 3; 65pp; English.
XX
```

CC   The present invention relates to nucleic acid that encodes a rapidly
CC   maturing chromo or fluorescent mutant of a Cnidarian chromo- or
CC   fluorescent protein or its mutant. The protein is useful in applications
CC   involving nucleic acid encoding a chromo- or fluorescent protein and is
CC   useful for producing a chromo and/or fluorescent protein which involves
CC   growing the cell, whereby the protein is expressed, and isolating the
CC   protein substantially free of other proteins. The protein and is useful in
CC   applications involving chromo- or fluorescent protein and is useful as
CC   PCR primers, hybridization probes, etc. The expression cassettes are useful
CC   for synthesizing related proteins. The chromoproteins are useful
CC   as coloring agents which are capable of imparting color or pigment to a
CC   particular composition of matter e.g. food compositions, pharmaceuticals,
CC   cosmetics, living organisms, e.g., animals and plants. The chromoproteins
CC   may also find use as labels in analyte detection assays, e.g. assays for
CC   biological analytes of interest and as selectable markers in recombinant
CC   DNA applications, e.g. the production of transgenic cells and organisms.
CC   The fluorescent proteins find use in a variety of different applications,
CC   e.g. in fluorescence resonance energy transfer (FRET) applications, as
CC   bioensors in prokaryotic and eukaryotic cells, in applications involving
CC   the automated screening of arrays of cells expressing fluorescent
CC   reporting groups by using microscopic imaging and electronic analysis, as
CC   second messenger detectors, and in fluorescence activated cell sorting
CC   applications and as in vivo marker in animals. The fluorescent proteins
CC   also find use in protease cleavage assays. The proteins can also be used
CC   is assays to determine the phospholipid composition in biological
CC   membranes and as a fluorescent timer. The present sequence represents the
CC   mutant-type DsRED encoding sequence.
XX
SQ   Sequence 704 BP; 145 A; 237 C; 212 G; 110 T; 0 U; 0 Other;

Alignment Scores:
```
Pred. No.:                 1.19e-133
Score:                     1160.00      Length:        704
Percent Similarity:        97.33%       Matches:       216
Best Local Similarity:     96.00%       Conservative:    3
Query Match:               95.55%       Mismatches:      6
DB:                        11           Indels:          0
                                        Gaps:            0
```

US-10-006-922A-12 (1-225) x ADN33980 (1-704)

```
Qy    1 MetArgSerSerLysAsnValIleLeuValSerGluPheMetArgPheLysValArgMetGluGly 20
        |||      | |||| ||:::|||||||||||||||||||||||||||||||||||||||
Db   20 ATGGCGTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGGATGGAGGGC 79

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   80 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 139

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  140 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 199

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  200 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 259

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  260 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 319

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  320 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 379

Qy  121 LysValLysPheIleGlyValAsnPheProSerArgGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  380 AAGGTGAAGTTCATCGGCGTGAACTTCCCGTCCCGAGGCCCCGTAATGCAGAAGAAGACT 439

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  440 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 499

Qy  161 IleHisLysAlaLeuLysLysAspGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  500 ATCCACAAGGCCCTGAAGAAGGACGGGCACTACCTGGTGGAGTTCAAGTCCATC 559

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  560 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 619

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  620 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 679

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  680 CACCACCTGTTCCTG 694
```

```
RESULT 69

ID   ADM13536 standard; DNA; 678 BP.
XX
AC   ADM13536;
XX
DT   15-JUL-2004  (first entry)
XX
DE   DNA encoding superfolding coral DsRed protein #1.
XX
KW   ADM13536;
KW   GFP; superfolder mutant; ds; green fluorescent protein;
KW   directed-evolution; folding interference domain; reporter protein;
KW   mutant; DsRed; coral.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
PN   US2004078148-A1.
XX
PD   22-APR-2004.
XX
PF   24-APR-2003; 2003US-0042368.
XX
PR   24-APR-2002; 2002US-00132067.
XX
PA   (WALD/) WALDO G S.
PA
PI   Waldo GS;
XX
DR   WPI; 2004-340059/31.
DR   P-PSDB; ADM13537.
XX
PT   New modified green fluorescent protein, useful as a reporter in
PT   expression studies.
XX
PS   Disclosure; SEQ ID NO 40; 46pp; English.
XX
CC   The invention relates to a green fluorescent protein (GFP) that comprises
CC   at least 80% identity to the GFP superfolding mutant appearing as
CC   ADM13535 that has at least one amino acid substitution selected from
CC   e.g., a substitution at position 30 that is an arginine or a conservative
CC   variant of arginine and measurable fluorescence activity. Also included
CC   are directed-evolution for generating an enhanced folding variant of a
CC   target polypeptide (comprising mutating a polynucleotide encoding
CC   polypeptide of interest to generate a library of mutated polynucleotides,
CC   linking mutated polynucleotides to polynucleotide encoding folding
CC   interference domain to form fusion protein (FP) constructs, expressing
CC   FP, and selecting FP that display optimal folding activity in relation to
CC   FP comprising wild-type polypeptide and folding interference domain, thus
CC   identifying polypeptide with enhanced folding activity), an enhanced
CC   folding variant of a fluorescent protein generated by the method, an
CC   enhanced folding variant of a chromophoric protein generated by the
CC   method, enhancing folding of a polypeptide comprising multiple domains
CC   (involving joining a first domain of the polypeptide to a poorly folding
CC   FP, and selecting FP that display optimal folding activity in relation to
CC   increase in the amount of activity generated by a first mutated fusion
CC   protein in comparison to a fusion protein comprising a wild-type first
CC   domain and the poorly folding polypeptide domain, thus identifying a
CC   first domain with enhanced folding, joining a second domain of the
CC   polypeptide to the first mutated fusion protein to form a second fusion
CC   protein, mutating the second domain, and detecting an increase in the
CC   amount of activity generated by a second mutated fusion protein in
CC   comparison to a fusion protein comprising the wild-type second domain and
CC   the first mutated fusion protein, thus identifying a target polypeptide
CC   with multiple domains that have enhanced folding), an isolated nucleic
CC   acid encoding the GFP , an expression vector comprising the nucleic acid
CC   and a host cell comprising the vector. The DsRed protein from coral is
CC   also considered for mutation using the above method. The enhanced folding
CC   variants are useful as reporter proteins to express the report level of a
CC   protein. The method is useful for directed-evolution of generating
CC   enhanced folding variant of target polypeptide and for improving folding
CC   and solubility of a target protein. The present sequence encodes a coral
```

```
CC   DsRed enhanced folding variant. NOTE: The authors refer to SEQ ID 38-42
CC   as SEQ ID 1-5 throughout the specification and never refer to SEQ ID 1-37
CC   by number, it is obvious that SEQ ID 38-42 are meant to be SEQ ID 1-5.
XX
SQ   Sequence 678 BP; 203 A; 127 C; 181 G; 167 T; 0 U; 0 Other;
```

US-10-006-922A-12 (1-225) x ADM13536 (1-678)

Alignment Scores:
Pred. No.:   3.32e-130
Score: 1132.00
Percent Similarity: 95.56%
Best Local Similarity: 92.44%
Query Match: 93.25%

	Length:	678
	Matches:	208
	Conservative:	7
	Mismatches:	10
	Indels:	0
	Gaps:	0

```
Qy        1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        1 ATGGAGTCTTCCGACGGATGTTATCAAGGAGTTCATGAGAGTTTAAGGTTCACATGAGAGGA   60

Qy       21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 TCGGTCAAATGGGCACCAGAGTTTGAAATAGAAGGCGAAGGCGAAGGCCATACGGAAGGC  120

Qy       41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ACCCAGAACGTTAAAGCTTAAGGTAACTAAGGGGGGACCACCTTTGCCATTTGCTTGGGATATT  180

Qy       61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||::
Db      181 TTGTCACCACAATTTCAGTATGGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy       81 AspThrTyrLysLeuLysPheProGluGlyPhelysTrpGluArgValMetAsnPheGlu  100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAAATGGAAGAGGTCATGAACTTTGAA  300

Qy      101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
              |||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db      301 GACGGTGGCGTCGCTACTGTAACCCAGGATTCCAGTTTGGGAGATGGCTGTTTGATCTAC  360

Qy      121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 AAGGTCAAAGTTCAATGGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  420

Qy      141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
              |||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db      421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAT  480

Qy      161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db      481 ATTCATAAAGGCTCTGAAGCTGAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  540

Qy      181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600

Qy      201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 ATAACAAACCACCAAACGAAGACTATACAAATCGTTGAGCAGTATGAAAGAGCCGAGGGACGC  660

Qy      221 HisHisHisLeuPheLeu  225
              ||||||||||||||||||
Db      661 CACCACTGTTCCTT  675
```

```
RESULT 70

ID   ADC24130 standard; DNA; 681 BP.
XX
AC   ADC24130;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Discosoma red fluorescent protein variant dimer2 DNA.
```

```
XX
KW   Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW   fluorescence protein variant; transcription protein detection;
KW   fluorescence energy resonance transfer; FRET; protein kinase;
KW   protein phosphatase; ion indicator; mutant; dimer2; ds.
XX
OS   Synthetic.
OS   Discosoma.
XX
PN   US20030059835-A1.
XX
PA   (TSIE/) TSIEN R Y.
PA   (CAMP/) CAMPBELL R E.
XX
PR   24-MAY-2001; 2001US-00866538.
PR   26-FEB-2001; 2001US-00794308.
PD   10-APR-2002; 2002US-00121258.
XX
PD   27-MAR-2003.
XX
PI   Tsien RY, Campbell RB;
XX
DR   WPI; 2003-743764/70.
DR   P-PSDB; ADC24131.
XX
PS   Example 1; SEQ ID NO 5; 67pp; English.
XX
PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.
XX
CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   containing a vector which comprises (I) operatively linked to an.
CC   expression control sequence, and an unit to assay the variant fluorescent
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcription, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This sequence encodes Discosoma red
CC   fluorescent protein variant dimer2.
XX
SQ   Sequence 681 BP; 145 A; 231 C; 205 G; 100 T; 0 U; 0 Other;
```

Alignment Scores:

```
Pred. No.:

Score:                 1121.00
Percent Similarity:    95.96%
Best Local Similarity: 92.34%
Query Match:           10
DB:
```

```
US-10-006-922A-12  (1-225) x ADC24130 (1-681)
```

```
Length:        681
Matches:       207
Conservative:  7
Mismatches:    9
Indels:        0
Gaps:          0
```

```
Qy     3  SerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyThrVal   22
          |||||||:::||||||||||||||||||||||||||||||||||||||||||||:::||
Db    10  TCCTCCAAGAACGTCATCAACAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTG   69

Qy    23  AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn   42
          ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db    70  AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAAC   129

Qy    43  ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer   62
          ||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db   130  ACCGTCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC   189

Qy    63  ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr   82
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   190  CCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTAC   249

Qy    83  LysLysLeuSerPheProGluGlyPhePheLysTrpGluArgValMetAsnPheGluAsp   102
          ||||||||||||||||||||||||||||:||||||||||||||||||:||
Db   250  AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC   309

Qy   103  GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal   122
          |||||||||||||||||||||||||||||||||||||||||||||:||||||
Db   310  GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG   369

Qy   123  LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly   142
          ||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db   370  AAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC   429

Qy   143  TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis   162
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   430  TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC   489

Qy   163  LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet   182
          ::|||||||||||||||||||||||||||:::||||||||||||||||||||||
Db   490  CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATG   549

Qy   183  AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr   202
          |||||||||||||||||||||||||||||||||||||:::||||||||
Db   550  GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACC   609

Qy   203  SerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHis   222
          ||||||||||||||||||||||||||||:::||||||||||||||||
Db   610  TCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGCCACCAC   669

Qy   223  LeuPheLeu   225
          |||||||
Db   670  CTGTTCCTG   678
```

```
XX
PN   WO200308446-A1.
XX
PD   23-OCT-2003.
XX
```

RESULT 71
ADL46209
ID   ADL46209 standard; DNA; 681 BP.
XX
AC   ADL46209;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Discosoma red fluorescent protein (DsRed) variant dimer2 coding sequence.
XX
KW   ds; gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay.
XX
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..681
FT   /tag= a
FT   /product= "DsRed variant dimer2 protein"
XX

PP 09-APR-2003; 2003WO-US010879.
XX
PR 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-00209208.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Campbell RE, Baird GS;
XX
DR WPI; 2003-845265/78.
DR P-PSDB; ADL46208.
XX
PT New monomeric and dimeric Anthozoan fluorescent protein variants with
PT reduced propensity to oligomerize, and encoding polynucleotides, useful
PT in molecular biology, e.g. in immunoassays or in tracking protein
PT movement in cells.
XX
PS Disclosure; SEQ ID NO 7; 166pp; English.
XX
CC The invention relates to a polynucleotide sequence encoding a Discosoma
CC red fluorescent protein (DsRed) variant having a reduced propensity to
CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC sequence, where the substitutions result in reduced propensity of the
CC DsRed variant to form tetramers and where the variant displays detectable
CC fluorescence at at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the variant
CC coding sequence for the variant dimer2.
XX
SQ Sequence 681 BP; 145 A; 231 C; 205 G; 100 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:　　　　　　　3
Score:　　　　　　　1121.00
Percent Similarity:　　　95.96%
Percent Identity:　　　92.83%
Best Local Similarity:　92.34%
Query Match:　　　　11

DB:
US-10-006-922A-12 (1-225) x ADL46209 (1-681)

Qy　　3 SerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
　　　　||||||||||||:||||:||||||||||||||||||||||:|||||
Db　10 TCCTCGAGGAGGTCATCAAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTG 69

Qy　23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
　　　　|||||||||||||||||||||||||||||||||||||||||||
Db　70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAAG 129

Qy　43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
　　　　|||||||||||||||||||||||||||||||||||||||||||
Db　130 ACCGGCCAAGGTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 189

Qy　63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
　　　　||||||||||||||||||||||||||||||||:||||:|||||
Db　190 CCCCAGTTCCAGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGACTAC 249

Qy　83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
　　　　||||||||||||||||||||||||||||||||||||||||||
Db　250 AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 309

Qy　103 GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal 122
　　　　|||||||||||||||||||||||||||||||||||||||||
Db　310 GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG 369

Qy　123 LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
　　　　|||||||||||||||||||||||||||||||||||||||||
Db　370 AAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACCATGGGC 429

Qy　143 TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis 162
　　　　|||||||||||||||||||||||||||||||||||||||||
Db　430 TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC 489

Qy　163 LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet 182
　　　　:||||||||||||||||||||||||||||||||:|||||
Db　490 CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATG 549

Qy　183 AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr 202
　　　　|||||||||||||||||||||||||||||||||||||||||
Db　550 GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACC 609

Qy　203 SerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHis 222
　　　　|||||||||||||||||||||||||||||:|||||||||||
Db　610 TCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGCCACCAC 669

Qy　223 LeuPheLeu 225
　　　　|||||||||
Db　670 CTGTTCCTG 678

RESULT 72
AAA48748
ID AAA48748 standard; cDNA; 678 BP.
XX
AC AAA48748;
XX
DT 19-SEP-2000 (first entry)
XX
DE Humanised anthozoa novel fluorescent protein drFP583/dmFP592 hybrid cDNA.
XX
KW Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW fluorescent protein; hybrid; 8S.
XX
OS Discosoma sp; "red".
OS Synthetic.
XX
FH Key　　　Location/Qualifiers
FT CDS　　　1..678
FT 　　　　　/product= "drFP583/dmFP592 hybrid protein"
XX
PN WO200034326-A1.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US029473.
XX
PR 11-DEC-1998; 98US-00210330.
PR 14-OCT-1999; 99US-00418529.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI Ding L;
XX
DR WPI; 2000-423381/36.
DR P-PSDB; AAY99838.
XX
PT Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT useful for fluorescent labeling and as markers.
XX
PS Claim 6; Page 77; 86pp; English.
XX
CC The present sequence is the nucleotide sequence of humanised
CC drFP583/dmFP592. Amplified wild-type coding region fragments from drFP583
CC and dmFP592 were digested with restriction endonucleases and combined
CC together to produce a hybrid construct. drFP583 is a full-length cDNA
CC encoding a novel fluorescent protein (nFP) from Discosoma sp. "red", a
CC non-bioluminescent species of the Class Anthozoa. The hybrid sequence was
CC changed to optimise the codons for expression of the protein in mammalian
CC cells. Fluorescent proteins can be used in fluorescent labeling, a useful
CC tool for marking a protein, cell or organism of interest. Unlike other
CC markers used in protein labeling, such as beta-galactosidase and
CC luciferase, fluorescent proteins do not require an exogenous cofactor or

CC substrate. Methods involving fluorescent proteins are also less laborious
CC and less difficult to control than the traditional methods of fluorescent
CC labeling, where a protein of interest is purified and then covalently
CC conjugated to a fluorophore derivative. Novel fluorescent proteins
CC isolated from species of the Class Anthozoa can be used as markers for
CC gene expression and protein localization studies, and in fluorescence
CC resonance energy transfer (FRET) reactions. They may have improved
CC properties and better suitability for larger excitations compared to
CC prior art fluorescent proteins such as green fluorescent protein
XX
SQ Sequence 678 BP; 145 A; 217 C; 218 G; 98 T; 0 U; 0 Other;

US-10-006-922a-12 (1-225) x AAA48748 (1-678)

Alignment Scores:
Pred. No.:
Score: 1.36e-128
Percent Similarity: 1119.00
Best Local Similarity: 96.89%
Query Match: 91.56%
DB: 92.17%
                         3

Length: 678
Matches: 206
Conservative: 12
Mismatches: 7
Indels: 0
Gaps: 0

```
Qy    1 MetArgSerSerLysSerAsnValIleLeuLysGlyIlePheMetArgPheLysValArgMetGluGly  20
Db    1 ATGAGCTGCAGCAAGAACGTGATCAAGGAGTTCATCCGGTTCAAGGTGCGGATGGAGGGC  60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  121 CACAACACCGTGAAGCTGAAGGTCACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  181 CTCAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCTGCTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAGGCCAGCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACAGCAAGCTCGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGG 660

Qy  221 HisHisHisLeuPheLeu 225
Db  661 CACCACCTCTTCCTC 675
```

RESULT 73
AAD46280
ID   AAD46280 standard; DNA; 678 BP.
XX
AC   AAD46280;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Discosoma sp. drFP583/dmFP592 (6/9Q) protein encoding DNA.
XX
KW   Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW   fluorescence activated cell sorting application; fluorescent timer;
KW   biosensor; fluorescence resonance energy transfer application; FRET;
KW   colouring agent; recombinant DNA application; analyte detection assay;
KW   sunscreen; second messenger detector; drFP583 protein; dmFP592 protein;
KW   gene; ds.
XX
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /*tag= a
FT   /product= "drFP583/dmFP592 (6/9Q) protein"
XX
PN   WO200268459-A2.
XX
PD   06-SEP-2002.
XX
PF   20-FEB-2002; 2002WO-US005749.
XX
PR   21-FEB-2001; 2001US-0270983P.
PR   04-DEC-2001; 2001US-00006922.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov S, Lukyanov K, Yanushevich Y, Savistky A, Fradkov A;
XX
DR   WPI; 2002-691654/74.
DR   P-PSDB; AAE28835.
XX
PT   New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT   of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
PT   analyte detection assays or fluorescence activated cell sorting
PT   applications.
XX
PS   Disclosure; Page 72; 80pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding non-aggregating
CC   chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC   useful in analyte detection assays, as colouring agents, as markers in
CC   recombinant DNA applications, as sunscreens or filters, in fluorescence
CC   resonance energy transfer (FRET) applications, as biosensors in
CC   prokaryotic and eukaryotic cells, in screening assays, as second
CC   messenger detectors, in protease cleavage assays or as fluorescent timers,
CC   in fluorescence activated cell sorting applications, The present
CC   sequence is a DNA encoding Discosoma sp. drFP583/dmFP592 (6/9Q) protein
CC   of the invention
XX
SQ   Sequence 678 BP; 145 A; 217 C; 218 G; 98 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AAD46280 (1-678)

Alignment Scores:
Pred. No.:
Score: 1.36e-128
Percent Similarity: 1119.00
Best Local Similarity: 96.89%
Query Match: 91.56%
DB: 92.17%
                         6

Length: 678
Matches: 206
Conservative: 12
Mismatches: 7
Indels: 0
Gaps: 0

```
Qy    1 MetArgSerSerLysSerAsnValIleLeuLysGlyIlePheMetArgPheLysValArgMetGluGly  20
Db    1 ATGAGCTGCAGCAAGAACGTGATCAAGGAGTTCATCCGGTTCAAGGTGCGGATGGAGGGC  60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 60
```

RESULT 74
AAA48749
ID   AAA48749 standard; cDNA; 678 BP.
XX
AC   AAA48749;
XX
DT   19-SEP-2000 (first entry)
XX
DE   Humanised Anthozoa novel fluorescent protein drFP583/dmFP592-2G.
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent labeling; hybrid; mutant; mutein; ss.
XX
OS   Discosoma sp; "red".
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /product= "drFP583/dmFP592-2G mutant hybrid protein"
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029473.
XX
PR   11-DEC-1998; 98US-00210330.
PR   14-OCT-1999; 99US-00418529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX

PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;
PI   Ding L;
XX
DR   WPI; 2000-423381/36.
DR   P-PSDB; AAY99839.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 6; Page 79; 86pp; English.

XX
CC   The present sequence is derived from the nucleotide sequence of humanised
CC   hybrid drFP583/dmFP592. drFP583 is a full-length cDNA encoding a novel
CC   fluorescent protein (nFP) from Discosoma sp. "red", a non-bioluminescent
CC   species of the Class Anthozoa. The hybrid sequence was changed to
CC   optimise the codons for expression of the protein in mammalian cells. The
CC   present sequence encodes two substitutions, K15Q and T217S, which were
CC   incorporated by random mutagenesis. Fluorescent proteins can be used in
CC   fluorescent labeling, a useful tool for marking a protein, cell or
CC   organism of interest. Unlike other markers used in protein labeling, such
CC   as beta-galactosidase and luciferase, fluorescent proteins do not require
CC   an exogenous cofactor or substrate. Methods involving fluorescent
CC   proteins are also less laborious and less difficult to control than the
CC   traditional methods of fluorescent labeling, where a protein of interest
CC   is purified and then covalently conjugated to a fluorophore derivative.
CC   Novel fluorescent proteins isolated from species of the Class Anthozoa
CC   can be used as markers for gene expression and protein localization
CC   studies, and in fluorescence resonance energy transfer (FRET) reactions.
CC   They may have improved properties and better suitability for larger
CC   excitations compared to prior art fluorescent proteins such as green
CC   fluorescent protein
XX
SQ   Sequence 678 BP; 143 A; 218 C; 218 G; 99 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AAA48749 (1-678)

Alignment Scores:
Pred. No.:   1.33e-127   Length: 678
Score: 1111.00   Matches: 204
Percent Similarity: 96.89%   Conservative: 14
Best Local Similarity: 96.67%   Mismatches: 7
Query Match: 91.52%   Indels: 0
DB: 3   Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuGluValPheMetArgPheLysValArgMetGluGlyGly 20
         |||
Db   1 ATGAGGTCGAGCAAGAACGTGATCAAGGAGGTCATGCGGTTCAAGGTGCGGATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||:::||||||:::||||||:||:||||:||||:||:||:||||:::||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAAGGCGAAGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||
Db 121 CACTGCAGCCTGAAACTGAAGGTGACCAAGGGCGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db 181 CTCAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAAGCTCAGCTTCCCCGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTCCAAGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 GAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db   61 ACCGTGAACGGCCACGAGTTCGAGATCAAGGGCGTAGGGCCGAGGGGCGGAGGGGCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||
Db 121 CACTGCAGCCTGAGCCTCATGGTGACCAAGGGCGGCGGCCCCCTTCGGCTTCGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysTyrProAlaAspIlePro 80
         |||
Db 181 CTCAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAAGCTCAGCTTCCCCGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTCCAAGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 GAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 CGGGGCGGCTGGGAGGCCAGCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGGAC 480

Qy 161 IleHisGlyAlaLeuLysLeuAlaAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCACATGGCCCTCCGGGCTCGAGGCGGCGCCCACTACCTCGTGGAGTTCAAGAGCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTTCCCGGGCTACTACTACGTGGACAGCAAGCTCGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACAGAGGGCCG 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACCTCTTCCTC 675

RESULT 75

ID   AAA48750 standard; cDNA; 678 BP.
XX
AC   AAA48750;
XX
DT   19-SEP-2000 (first entry)
XX
DE   Humanised Anthozoa novel fluorescent protein drFP583/dmFP592-Q3 cDNA.
XX
KW   Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW   fluorescent; labeling; hybrid; mutant; mutein; ss.
XX
OS   Discosoma sp; "red".
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..678
FT   /product= "drFP583/dmFP592-Q3 mutant hybrid protein"
XX
PN   WO200034326-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US029473.
XX
PR   11-DEC-1998;   98US-00210330.
PR   14-OCT-1999;   99US-00418529.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA,   Fradkov AF,   Labas YA,   Matz MV,   Green G,   Chen Y;
PI   Ding L;
XX
DR   WPI; 2000-423381/36.
DR   P-PSDB/ AAY99840.
XX
PT   Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT   useful for fluorescent labeling and as markers.
XX
PS   Claim 6; Page 80-81; 86pp; English.
XX
CC   The present sequence is derived from the nucleotide sequence of humanised
CC   hybrid drFP583/dmFP592. drFP583 is a full-length cDNA encoding a novel
CC   fluorescent protein (nFP) from Discosoma sp. "red", a non-bioluminescent
CC   species of the Class Anthozoa. The hybrid sequence was prepared to
CC   optimise the codons for expression of the protein in mammalian cells. The
CC   present sequence encodes three substitutions, K15Q, K83M and T217S, which
CC   were incorporated by random mutagenesis. Fluorescent proteins can be used
CC   in fluorescent labeling, a useful tool for marking a protein, cell or
CC   organism of interest. Unlike other markers used in protein labeling, such
CC   as beta-galactosidase and luciferase, fluorescent proteins do not require
CC   an exogenous cofactor or substrate. Methods involving fluorescent
CC   proteins are also less laborious and less difficult to control than the
CC   traditional methods of fluorescent labeling, where a protein of interest
CC   is purified and then covalently conjugated to a fluorophore derivative.
CC   Novel fluorescent proteins isolated from species of the Class Anthozoa
CC   can be used as markers for gene expression and protein localization
CC   studies, and in fluorescence resonance energy transfer (FRET) reactions.
CC   They may have improved properties and better suitability for larger
CC   excitations compared to prior art fluorescent proteins such as green
CC   fluorescent protein

Alignment Scores:
Pred. No.:
Score: 7.36e-127
Percent Similarity: 1105.00
Best Local Similarity: 96.44%
Query Match: 90.22%
DB: 91.02%
                      3

SQ   Sequence 678 BP; 142 A; 218 C; 218 G; 100 T; 0 U; 0 Other;

US-10-006-922A-12 (1-225) x AAA48750 (1-678)

Qy   1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGAGCTCACGAAGAACGTGATCAAGGAGTTCATGCGGTTCAAGGTGCGGATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAAGGCGAGGGCGAGGGCCGGCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   121 CACTGCAACGTGAAGCTCATCGTGACCAAGGGCGGCCCCCTGCCCTTCGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   181 CTCAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPhePheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   241 GACTACAAGAAGCTCAGCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   421 ATGGGCTGGGAGGCCAGCACGGAGCGCCTCTACCCCCGGGACGGCGTGCTCAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACAGCAAGCTCGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   601 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGG 660

Qy   221 HisHisLeuPheLeu 225
Db   661 CACCACCTCTTCCTC 675

RESULT 76

ID   AAA28958 standard; cDNA; 876 BP.
XX

AC   AAA28958;
XX
DT   12-SEP-2000 (first entry)
XX
DE   Discosoma sp. "magenta" dmFP592 cDNA encoding a fluorescent protein.
XX
KW   dmFP592; fluorescent protein; non-bioluminescent; label; marker;
KW   fluorescence resonance energy transfer reaction; FRET; ss.
XX
OS   Discosoma sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   45..737
FT   /tag= a
FT   /product= "fluorescent_protein"
XX
PN   WO200034324-A1.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999; 99WO-US029412.
PR   11-DEC-1998; 98US-00210330.
PR   19-NOV-1999; 99US-00443341.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
XX
DR   WPI; 2000-423379/36.
DR   P-PSDB; AAY92894.
XX
CC   Novel fluorescent protein from non-bioluminescent Discosoma species,
CC   useful for fluorescent labeling and as markers.
XX
PS   Claim 7; Page 54-55; 60pp; English.
XX
CC   This cDNA encodes a fluorescent protein (dmFP592) from Discosoma sp.
CC   "magenta", a non-bioluminescent species of the genera Anthozoa. The
CC   fluorescent colour is observed under usual white light or ultraviolet
CC   light. It has an absorbance maximum at 573 nm and an emission maximum at
CC   593 nm. The polynucleotides of the invention can be used to produce the
CC   proteins recombinantly, and as a source of primers and probes for
CC   identifying related proteins. The fluorescent proteins have applications
CC   in fluorescent labeling, as fluorescent markers for gene expression and
CC   protein localization studies, and in fluorescence resonance energy
CC   transfer (FRET) reactions. They may have improved properties and better
CC   suitability for larger excitations, compared to prior art fluorescent
CC   proteins such as green fluorescent protein from Aequorea victoria, which
CC   has an excitation maximum at 395 nm, a second excitation peak at 475 nm
CC   and an emission maximum at 510 nm.
XX
SQ   Sequence 876 BP; 256 A; 165 C; 232 G; 223 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.75e-124
                                1085.50
Percent Similarity: 95.58%
                                88.05%
Best Local Similarity: 89.42%
Query Match: 3

Length: 876
Matches: 199
Conservative: 17
Mismatches: 9
Indels: 1
Gaps: 1

US-10-006-922A-12 (1-225) x AAA28958 (1-876)

Qy   1 MetArgSerSerLeuLysGluAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
            ||| ||||||| || ||| || ||||| ||| |||
Db   45 ATGAGTTGTTCCAAGAAATGTGATCAAGGAGTTCATCGAGTTCAAGGTTCGTATGGAAGGA 104

Qy   21 ThrValAsnGlyHisGlyPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db 105 ACGGTCAATGGGCACGAGTTTGAAATAAAAGGCGAAGGTGAAGGGGAGGGCCTTACGAAGGT 164

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db ...

Db 165 CACTGTTCCGTAAAGCTTATGGTGTAACCAAGGGTGGA... 224
Db 225 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAAACACCCTGCCGACAATACCA 284
Db 285 GACTATAAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGAAAAGGGTCATGAACTTTGAA 344
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 345 GACGCGGTGGCCGTGGTTACTGTATCCCAAGATTCCAGTTTGAAAGACGGGCCGTTTCATCTAC 404
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 405 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCCGTGATGCAGAAGAGGACA 464
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysArgThr 140
Db 465 CGGGGCTGGGAAGCCAGCACCGAGATGCTGTATCCTGCCGACGGTGGCGTGCTGACAGCGAC 524
Qy 141 MetGlyTrpGluAlaSerThrGluMetLeuTyrProAlaAspGlyGlyLeuGluGlyGlu 160
Db 525 ATCCATATGGCCTCTGAGGCTGGAAGGAGGGGCCATTACCTCGTTGAAATTCAAAAGTATT 584
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 585 TACATGGTAAAGAGAGCCTTCAGTGCGACTCTATTAATGTTGACTCCAAACTG 644
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrTyrProValAlaAspSerLeu 199
Db 645 GATATGACGAGCCACCAACGAAGATTACACAGTCGTTGAGCAGTATGAAAAAACCCAGGGA 704
Qy 200 AspIleThrSerHisAsnGluAspTyrThrIleValGluGlnInTyrGluArgThrGluGlyInTyrGluGly 219
Db 705 CGCCACCATCCGTTCATT 722
Qy 220 ArgHisHisLeuPheLeu 225

RESULT 77
AAD03618
ID   AAD03618 standard; cDNA; 876 BP.
XX
AC   AAD03618;
XX
DT   11-SEP-2003 (revised)
DT   19-JUN-2001 (first entry)
XX
DE   Discosoma sp. magenta chromo/fluorescent protein, dmFP592 (NFP-9) cDNA.
XX
KW   Anthozoa; Chromoprotein; fluorescent protein; dmFP592; NFP-9; sunscreen;
KW   analyte detection assay; selectable marker; recombinant DNA application;
KW   biosensor; pH indicator; invivo marker; selective filter; ss.
XX
OS   Discosoma sp; magenta.
XX
PN   WO200127150-A2.
XX
PD   19-APR-2001.
XX
PF   13-OCT-2000; 2000WO-US028477.
XX
FH   Key   Location/Qualifiers
FT   CDS   45..737
FT   /tag= a
FT   /product= "Chromo/fluorescent protein, dmFP592 (NFP-9)"
XX
PR   14-OCT-1999; 99US-00418529.
PR   15-OCT-1999; 99US-00418917.
PR   15-OCT-1999; 99US-00418922.
PR   19-NOV-1999; 99US-00444338.
PR   19-NOV-1999; 99US-00444341.
PR   09-DEC-1999; 99US-00457556.

```
PR   09-DEC-1999;
PR   09-DEC-1999;  99US-00457898.
PR   09-DEC-1999;  99US-00458144.
PR   09-DEC-1999;  99US-00458477.
PR   10-DEC-1999;  99WO-US029405.
PR   14-JUN-2000;  2000US-02116070P.
PR   14-JUN-2000;  2000US-02116209P.
PR   14-JUN-2000;  2000US-02116260P.
PR   14-JUN-2000;  2000US-02116270P.
PR   14-JUN-2000;  2000US-02116566P.
PR   14-JUN-2000;  2000US-02116627P.
PR   14-JUN-2000;  2000US-02116762P.
PR   14-JUN-2000;  2000US-02117566P.
PR   14-JUN-2000;  2000US-02117667P.
PR   14-JUN-2000;  2000US-02117880P.
PR   14-JUN-2000;  2000US-02118880P.
PR   14-JUN-2000;  2000US-02118888P.
PR   14-JUN-2000;  2000US-02120070P.
XX
PA   (CLON-) CLONTECH LAB INC.
XX
PI   Lukyanov SA,  Fradkov AF,  Labas YA,  Matz MV,  Terskikh A;
XX
PN   WPI; 2001-266409/27.
DR   P-PSDB; AAE00378.
XX
PT   Claim 13; Fig 9; 69pp; English.

CC   The present sequence is a Discosoma sp. magenta chromo/fluorescent
CC   protein, dmFP592 (NFP-9) cDNA. NFP-9 is present in an environment other
CC   than its natural environment and has an absorbance maximum ranging from
CC   375nm to 775nm and more usually from 560nm to 590nm and emission maximum
CC   ranging from 395nm to 795nm and more usually from 580 to 610nm. The
CC   chromoproteins or fluorescent proteins are useful as labels in analyte
CC   detection assays, as selectable markers in recombinant DNA applications,
CC   as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
CC   and as in vivo markers in animals. They are also useful in sunscreens and
CC   as selective filters. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ   Sequence 876 BP; 256 A; 165 C; 232 G; 223 T; 0 U; 0 Other;
```

```
US-10-006-922A-12 (1-225) x AAD03618 (1-876)

Alignment Scores:
Pred. No.:
Score:                  2.75e-124
Score:                  1085.50
Percent Similarity:     95.58%
Best Local Similarity:  88.05%
Query Match:            89.42%
DB:                     4

                        Length:        876
                        Matches:       199
                        Conservative:  17
                        Mismatches:    9
                        Indels:        1
                        Gaps:          1
```

```
Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45  ATGAGTTGTTCCAAGAAATGTGATCAAGAGAGTTCATGAGGTTCGTATGGAGGAGGA  104

Qy   21  ThrValAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105  ACGGTCAATGGGCACGAGTTTGAAATAAAAGGCGAAGGTGAAGGGGAGGGCCTTACGGAGGT  164

Qy   41  HisAsnThrValLysThrLysGlyThrLysValLysThrLysGlyThrLysValLysPro  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  165  CACTGTTCCGTAAAGCTTATGGTAACCAAAGGGTGGGACCTTTGCCATTTGCTTTTGATTATT  224

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  225  TTGTCACCAATTTCAGTATGGAAGCAAAGGTATATGTCAAAACACCCGCACATACCA  284

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  285  GACTATAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAAGGGTCATGAACTTTGA  344
```

```
Db  345  GACGGTGGCGTGGGTTACTGTATCCAAGATTCCAGTTTGAAAGACGGCTGTTTCATCTAC  404

Qy  121  LysValGlyPheIleGlyValAsnProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  405  GAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCTGTTATGCAGAAGAAGACA  464

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         |||||||||||||||||||||||||||||||||||||::::::||||||||||||||||
Db  465  CGGGGCGCTGGGAAGCCAGCTCTGAGCGTTTGTATCCTCGTGATGGGGTGCTGAAAGGAGAC  524

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||::::::|||||||||||||||||||||||||||||||||||||||||||||||
Db  525  ATCCATAATGGCCTCTGAGGCTGGAAGGATGGGCGCCATTACCTCGTTGAAATTCAAAAGTATT  584

Qy  181  TyrMetAlaLysLysPro---ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeu  199
         |||||||||||||||||      |||||||||||||||||||||||||||||||||||||
Db  585  TACATGTGAAGAAGCCTTCAGTGCAGTTGCCAGGCTACTATTATGTTGACTCCAAACTG  644

Qy  200  AspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGly  219
         ||||||||||||||||||||||||||||||||||||||||||||||||::::||||||
Db  645  GATATGACGAGCCACAACGAAGAATTACACAGTCGTTGAGCAGTATGAAAAAACCCAGGA  704

Qy  220  ArgHisHisLeuPheLeu  225
         ||||||||||||:::
Db  705  CGCCACCACCTCCGTTCATT  722
```

```
Db   Novel eukaryotic chromosome comprising one or many att sites which
Db   permits site-directed integration in the presence of lambda-integrase,
Db   useful for site-specific recombination-directed integration of DNA of
Db   interest.
```

```
RESULT 78
ID   ACC44649
AC   ACC44649

ID   ACC44649 standard; DNA; 876 BP.
XX
AC   ACC44649;
XX
DT   29-MAY-2003  (first entry)
XX
DE   Red fluorescent protein (DsRed) encoding DNA SEQ ID NO:39.
XX
KW   Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
KW   att site; Integrase; recombinase; Aces; gene therapy; transgenic animal;
KW   platform artificial chromosome expression system; gene; ds.
XX
OS   Discosoma sp.
OS   Synthetic.
XX
PN   WO020297059-A2.
XX
PD   05-DEC-2002.
XX
DR   WPI; 2003-140461/13.
DR   P-PSDB; ABP96651.
XX
PI   Perkins E,  Perez C,  Lindenbaum M,  Greene A,  Leung J,  Fleming E;
PI   Stewart S,  Shellard J;
XX
PA   (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX
PR   30-MAY-2001; 2001US-0294758P.
PR   21-MAR-2002; 2002US-03668911P.
XX
PS   Example 1; Page 191-192; 272pp; English.
XX
CC   The present invention describes a eukaryotic chromosome (I) comprising
CC   one or several att sites, where an att site is heterologous to the
CC   chromosome, and permits site-directed integration in the presence of
CC   lambda-integrase. Also described: (I) a platform artificial chromosome
CC   expression system (ACes) (II) comprising several sites that participate
CC   in recombinase catalysed recombination; and (III) a method (M) for
```

Left column

```
CC   introducing a heterologous nucleic acid into a platform artificial
CC   chromosome. (I) can be used in gene therapy. (M) is useful for
CC   introducing a heterologous nucleic acid molecule into a platform
CC   artificial chromosome, preferably an ACes. (II) is useful for producing a
CC   transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
CC   mammal) by introducing (II) by cell fusion, lipid-mediated transfection,
CC   by a carrier system, microinjection, microcell fusion, electroporation,
CC   microprojectile bombardment or direct DNA transfer into an embryonic
CC   cell, preferably a stem cell or an embryo. (II) comprises a heterologous
CC   nucleic acid that encodes a therapeutic product which is useful for
CC   making a library of ACes comprising random portions of a genome. ACC44612
CC   to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 876 BP; 256 A; 165 C; 232 G; 223 T; 0 U; 0 Other;
```

US-10-006-922A-12 (1-225) x ACC44649 (1-876)

```
Alignment Scores:
Pred. No.:                  2.75e-124
Score:                      1085.50
Percent Similarity:         95.58%
Best Local Similarity:      88.05%
Query Match:                89.42%
DB:                         10

                 Length:        876
                 Matches:       199
                 Conservative:   17
                 Mismatches:      9
                 Indels:          1
                 Gaps:            1
```

```
Qy    1 MetArgSerSerLysGluAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
        ||||
Db   45 ATGAGTTGTTCCAAGGAAAATGTGATCAAGGAGTTCATGAGGTTCAAGGTTCGTATGGAAGGA  104

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
        |||||||
Db  105 ACGGTTGCAAATGGGCACGAGTTTGAAATAGAAGGGGAAGGTGAAGGGAGGCCTTACGAAGGT  164

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle      60
        |:||
Db  165 CACTGTTCCGTAAAGCTTATAGGTAACCAAGGGTGGACCTTTGCCATTTGCTTTTGATTATT   224

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro      80
        ||||
Db  225 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAACACCCTGCCGACATACCA    284

Qy   81 AspPheTyrLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
        ||||
Db  285 GACTTATAAAAAGGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  344

Qy  101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr    120
        ||||
Db  345 GACGGTGGCGTGGCTACTGTAACCCAAGATTCCAGTTTGAAAGACGGGCTGTTTCATCTAC  404

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
        ||||
Db  405 AAGGTCAAGTTCATAGGGGTGAACTTTCCTTCTGATGGACCCGTTATGCAGAAGAAGACA   464

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
        ||||
Db  465 CGGGGGCTGGAAGCCAGCTCTGAGCCGTTTGTATCCTCGTGATGGGGTGCTGAAGGAGAC   524

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
        |||||:||
Db  525 ATCCATATGGCTCTGAAGCTGGAAGACGGCGGCCATTACCTCGTTGAATTCAAAAGTATT   584

Qy  181 TyrMetAlaLysLysPro---ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeu   199
        |||||||                 |:|
Db  585 TACATGGTAAAGAAGCCTTCAGTCCAGCTACTAATTATGTGACTCCAAACTG           644

Qy  200 AspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGly   219
        |:|||
Db  645 GATATGACGAGCCACAACGAAGATTACACAGTCGTTGAGCAGTATGAAAAAACCAGGGA   704

Qy  220 ArgHisHisLeuPheLeu   225
        ||||:||
Db  705 CGCCACCACCATCCGTTCATT   722
```

Right column

```
RESULT 79
ADC24132
ID   ADC24132 standard; DNA; 678 BP.
XX
AC   ADC24132;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Discosoma red fluorescent protein variant mRFP1 DNA.
XX
KW   Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW   fluorescent protein variant; transcription induction detection;
KW   fluorescence energy resonance transfer; FRET; protein kinase;
KW   protein phosphatase; ion indicator; mutant; mRFP1; ds.
XX
OS   Synthetic.
OS   Discosoma.
XX
PN   US2003059835-A1.
XX
PD   27-MAR-2003.
XX
PF   10-APR-2002; 2002US-00121258.
XX
PR   26-FEB-2001; 2001US-00794308.
PR   24-MAY-2001; 2001US-0086538.
XX
PA   (TSIE/) TSIEN R Y.
     (CAMP/) CAMPBELL R E.
XX
PI   Tsien RY, Campbell RE;
XX
DR   WPI; 2003-743764/70.
DR   P-PSDB; ADC24133.
XX
PT   Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT   variant having a reduced propensity to oligomerize, useful for detecting
PT   transcriptional activity.
XX
PS   Disclosure; SEQ ID NO 7; 67pp; English.
XX
```

```
CC   The invention describes a polynucleotide sequence (I) encoding a
CC   Discosoma red fluorescent protein (DsRed) variant having a reduced
CC   propensity to oligomerise, comprising amino acid substitutions at the AB
CC   and/or AC interfaces of the wild-type DsRed sequence (SI) comprising 225
CC   amino acids, given in the specification, where the substitutions result
CC   in reduced propensity of the DsRed variant to form tetramers. (I) is
CC   useful for detecting transcriptional activity by providing a host cells
CC   containing a vector which comprises (I) operatively linked to an
CC   expression control sequence, and an unit to assay the variant fluorescent
CC   protein fluorescence, and assaying fluorescence of the variant
CC   fluorescent protein produced by (VII), where variant fluorescent protein
CC   fluorescence is indicative of transcriptional activity. A polynucleotide
CC   encoding a fusion protein is useful for the analysis of in vivo
CC   localisation or trafficking of a polypeptide of interest. A polypeptide
CC   marker is useful as markers to identify the location and amount of a
CC   target protein produced, where the target protein is fused to the marker,
CC   as a complement to or alternative for the green fluorescent protein or
CC   its spectral variant, for detecting induction of transcription, in
CC   applications involving fluorescence energy resonance transfer (FRET),
CC   which detects events as the function of the movement of fluorescent
CC   donors and acceptors towards or away from each other, for making
CC   fluorescent sensors for protein kinase and phosphatase activities or
CC   indicators for ions and molecules such as Ca2+, Zn2+, for identifying the
CC   presence of a molecule in a sample, for identifying a specific
CC   interaction of a first and second molecule, for determining whether a
CC   sample contains an enzyme or for determining the pH of the sample. (I) is
CC   useful for identifying a region or condition that regulates the activity
CC   of an expression control sequence. This sequence encodes Discosoma red
CC   fluorescent protein variant mRFP1.
XX
SQ   Sequence 678 BP; 149 A; 228 C; 208 G; 93 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:
Score:                  1.88e-116           Length:          678
Percent Similarity:     1021.00             Matches:         192
                        89.14%              Conservative:    5
Best Local Similarity:  86.88%              Mismatches:      24
Query Match:            84.10%              Indels:          0
DB:                     10                  Gaps:            0

US-10-006-922A-12 (1-225) x ADC24132 (1-678)

Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
Db    1  ATGGCCTCCTCCGAGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC   60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluIleGluGlyGluGlyArgProTyrGluGly  40
Db   61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  121  ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  181  CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  241  GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  361  AAGGTGAAGCGCGCCGGCGACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  421  ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGAGGAGGCGGGG  480

Qy  161  IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  481  ATCAAGATGAGGCTGAAGGCTGAAGGACGGGCGGCGCCACTACGGACGCGGAGGTCAAGACCACC  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACTACGTGGACTCCAAGCTGGAC  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluLeuTyrGlnArgThrGluGlyArg  220
Db  601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACTGTACCAGCGTACGGAGGGCCGC  660

Qy  221  His  221
Db  661  CAC  663
```

```
RESULT 80
ADL46211
ID   ADL46211 standard; DNA; 678 BP.
XX
AC   ADL46211;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Discosoma red fluorescent protein (DsRed)variant mRFP1 coding sequence.
XX
KW   red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay; gene;
KW   ds.
XX
OS   Discosoma sp.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..678
FT                  /tag= a
FT                  /product= "DsRed variant mRFP1 protein"
XX
PN   WO2003086446-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010879.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-00209208.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Tsien RY, Campbell RE, Baird GS;
XX
DR   WPI; 2003-845265/78.
DR   P-PSDB; ADL46210.
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PS   Disclosure; SEQ ID NO 9; 166bp; English.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   variant mRFP1 coding sequence.
XX
SQ   Sequence 678 BP; 149 A; 228 C; 208 G; 93 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:
Score:                  1.88e-116           Length:          678
Percent Similarity:     1021.00             Matches:         192
                        89.14%              Conservative:    5
Best Local Similarity:  86.88%              Mismatches:      24
Query Match:            84.10%              Indels:          0
DB:                     11                  Gaps:            0

US-10-006-922A-12 (1-225) x ADL46211 (1-678)

Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
Db    1  ATGGCCTCCTCCGAGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC   60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluIleGluGlyGluGlyArgProTyrGluGly  40
Db   61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  121  ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  181  CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
```

```
                        |||||
Qy  241  GACTACTTGAAGCTGTCCTTCCCCGAGGGGCTTCAAGTGTGGGAGGGCGGCGTGATGAACTTCGAG 300
         |||| |||||
Db  241  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyTyrCysPheIleTyr 300

Qy  101  IleHisLeuLysAlaLeuLeuLysGlyHisTyrGlyGlyHisIleTyrLeuValGluPheLysSerIle 120
Db  161  ...GACCGGGCCGGCTGGGTGACCGTGACCCCGAGCCAGACGTCCTCCCTGCAGGACGGCTACTGCTTCATCTAC 360

Qy  121  LysValValGlyPheIleLeuGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361  AAGGTGAAGCTGACGTGCGGGGACCAACTTCCCCTGCAGGCGGCCCCGTTAATGCAAGAAGAAGACC 420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421  ATGGGCTGGGAGGCGCCTCCACCGAGCGGGATGTACCCGCGGGACGGCGTCCCTGAAGGCGAG 480

Qy  161  IleHisLeuLysAlaLeuLeuLysGlyHisTyrGlyGlyHisIleTyrLeuValGluPheLysSerIle 180
Db  481  ATCAAGATGAGGCTGAAGGCTGAAGGACGGCGGCCACTACCGACGCGAGGTCAACACCACC 540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541  TACATGGCCAAGAAGCCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnThrGlyTyrHisGluProPheAlaTrpAspIle 220
Db  601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGACGGCGCCGCCGAGGGCCGC 660

Qy  221  His 221
Db  661  CAC 663
```

Alignment Scores:

Pred. No.:	1.01e-115	Length:	2241
Score:	1021.00	Matches:	192
Percent Similarity:	89.14%	Conservative:	5
Best Local Similarity:	86.88%	Mismatches:	24
Query Match:	84.10%	Indels:	0
		Gaps:	0
DB:	14		

US-10-006-922A-12 (1-225) x AEA54911 (1-2241)

```
Qy    1  MetArgSerSerLysAsnValIleIleLysGluHisMetArgPheLysValAlaArgMetGluGlyGly 20
         |||                |||::::|::|||::::|||
Db    1  ATGGCCTCCTCAGTTCCAGTTCCAAGGCCTTCACACCCCGGCCAGCATCCC 80

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ::::|||                :|||||||||||||||||
Db  805  ATGGGCCTCCTTCGGAGGACGTCATCAAGGAGGTTCATCGGCTTCAAGGTGTGGCCATGGAGGGC 864

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProProPheAlaTrpAspIle 60
         |||
Db  865  TCCGTCAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 924

Qy   61  LeuSerProGlnPheGlnTyrGlySerLeuValTyrValLysHisProAlaAspIlePro 80
         |||||                |||||::|||
Db  925  ACCCAGACCGCCACCAGCGTGAAGGTGACCAAGGGCGGCCCCCTGCCCCCTTCGCCTGGGACATC 984

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||                |||||::|||
Db  985  CTGTCCCCCTCCAGTTCCAGTTCCAAGGCCTCCAAGGACTACCGCTTCAAGGACATCCCC 1044

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyTyrCysPheIleTyr 120
Db 1045  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTACTGCTTCATCTAC 1104

Qy  121  LysValValGlyPheIleLeuGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 1105  AAGGTGAAGCTGACGTGCGGGGACGGCCCTAACTTCCCCTGCAGGCGGCCCCGTTAATGCAAGAAGAAGACC 1164

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 1165  ATGGGCTGGGAGGCGTCCACCGAGCGGCTGTACCCGCGGGACGGCGTGGCCCTGAAGGGCGAG 1224

Qy  161  IleHisLeuLysAlaLeuLeuLysGlyHisTyrGlyGlyHisIleTyrLeuValGluPheLysSerIle 180
Db 1225  ATCCACCTGAAGGCCCTGCTGAAGGGCCACTACGGCGGCGGCATCTACCCCGAGAAGGCCGAG 1284

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 1285  ATCAAGATGAGGCTGAAGGCTGAAGGACGGCGGCCACTACCGACGCGAGGTCAACACCACC 1344

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnThrGlyTyrHisGluProPheAlaTrpAspIle 220
Db 1345  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 1404

Qy  221  His 221
Db 1405  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGACGGCGCCGCCGAGGGCCGC 1464

Qy  221  His 221
Db 1465  CAC 1467
```

RESULT 82
ADQ59565

```
ID   ADQ59565 standard; DNA; 675 BP.
XX
AC   ADQ59565;
XX
DT   07-OCT-2004 (first entry)
XX
DE   Discosoma monomerised red fluorescent protein mRFP1 DNA.
XX
KW   optically detectable signal; protein fragment complementation assay; PCA;
KW   molecular interaction detection; biomolecular interaction detection;
KW   fluorescent protein; multi-colour PCA; drug discovery; target validation;
KW   high-throughput screening; high-content screening; pathway mapping;
KW   drug mechanism-of-action study; biosensor; diagnostic;
KW   red fluorescent protein; spectral shift; mRFP1; gene; ds.
XX
OS   Discosoma.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..675
FT                  /*tag= a
FT                  /product= "Discosoma monomerised red fluorescent protein"
XX
PN   US2004137528-A1.
XX
PD   15-JUL-2004.
XX
PF   01-DEC-2003; 2003US-0072417B.
XX
PR   02-FEB-1998; 98US-00017412.
PR   07-FEB-2000; 2000US-0499464.
PR   24-MAY-2002; 2002US-00154758.
PR   29-JAN-2003; 2003US-00353090.
PR   09-APR-2003; 2003US-0461133P.
XX
PA   (MICH/) WATSON MICHNICK S W.
PA   (MACD/) MACDONALD M L.
PA   (LAME/) LAMERDIN J.
XX
PI   Watson Michnick SW, Macdonald ML, Lamerdin J;
XX
PS   Disclosure; SEQ ID NO 15; 34pp; English.
XX
CC   The invention describes a composition (I) comprising complementary
CC   fragments of a protein or mutant protein, where the fragments generate an
CC   optically detectable signal when associated, and each of the mutant
CC   protein fragments is fused to a separate molecule. Also described are:
CC   protein fragment complementation assays (PCAs) for detection of molecular
CC   interactions, involving reassembling separate fragments from an optically
CC   detectable protein, and detecting the reassembly by units of
CC   reconstitution of activity of the optically detectable protein, where the
CC   reassembly of the fragments is operated by the interaction of other molecular
CC   domains fused to each fragment, and is independent of other molecular
CC   processes; detecting biomolecular interaction, involving selecting an
CC   appropriate optically detectable protein, effecting fragmentation of the
CC   optically detectable protein such that the fragmentation results in
CC   reversible loss of protein function, fusing or attaching fragments of the
CC   protein fragments separately to other molecules, reassociating
CC   the protein fragments through interactions of the molecules that are
CC   fused or attached to the fragments, and detecting the resulting optical
CC   signal; designing and engineering of PCAs based on fluorescent protein
CC   fragment complementation assays; and a method and composition for the
CC   construction of multi-color PCAs.
CC   (I) is useful in PCAs and other assays for drug discovery, target
CC   validation, high-throughput screening, high-content screening, pathway
CC   mapping, drug mechanism-of-action studies, biosensors and diagnostics.
CC   (I) is useful for engineering different colour PCAs for a variety of
CC   applications in biology and biotechnology. This sequence encodes a
CC   Discosoma monomerised red fluorescent protein from which PCA fragments
CC   can be isolated.
XX
SQ   Sequence 675 BP; 147 A; 229 C; 207 G; 92 T; 0 U; 0 Other;
```

US-10-006-922A-12 (1-225) x ADQ59565 (1-675)

Alignment Scores:

Pred. No.:	1	Length:	675
Score:	1015.00	Matches:	191
Percent Similarity:	88.69%	Conservative:	5
Best Local Similarity:	86.43%	Mismatches:	25
Query Match:	83.61%	Indels:	0
DB:	12	Gaps:	0

```
Qy  1   MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
        |||                    :::::::::::::::::::::::::::::
Db  1   ATGGCGTCCTCCAAGAAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
        |||||||||||||||||||||||||||||||||||||||||||
Db  61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy  41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
        |||
Db  121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy  81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
        |||||||||||||||||||||||||||||||||||||||||
Db  241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGA   300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
        |||||||||||||||||||||||||||||||
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
        |||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC  420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
        ||||||||||||||||||:::||||||||||||||||||
Db  421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCAGAGACGGCGTGCTGAAGGGCGAG  480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
        |||||||||||||||||||||||||||||||||||||||
Db  481 ATCCACAAGGCTCTGAAGCTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGAGCACC  540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
        |||||||||||||||||||||||||||||||
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC  600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
        |||||||||||||||||||||||||||||||||||||||
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGC  660

Qy  221 His  221
        |||
Db  661 CAC  663
```

```
RESULT 83
ADL46282
ID   ADL46282 standard; DNA; 678 BP.
XX
AC   ADL46282;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Discosoma red fluorescent protein (DsRed)variant mRFP1.1 coding sequence.
```

```
XX
KW   red fluorescent protein; DsRed; fluorescence; red wavelength;
KW   oligomerization; tetramerization; immunoassay; hybridization assay; gene;
KW   ds.
XX
OS   Discosoma sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..678
FT                   /*tag= a
FT                   /product= "DsRed variant mRFP1.1 protein"
XX
PN   WO200308646-A1.
XX
PD   23-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010879.
XX
PR   10-APR-2002; 2002US-00121258.
PR   29-JUL-2002; 2002US-00209208.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Tsien RY, Campbell RE, Baird GS;
XX
PT   New monomeric and dimeric Anthozoan fluorescent protein variants with
PT   reduced propensity to oligomerize, and encoding polynucleotides, useful
PT   in molecular biology, e.g. in immunoassays or in tracking protein
PT   movement in cells.
XX
PS   Claim 63; SEQ ID NO 80; 166pp; English.
XX
CC   The invention relates to a polynucleotide sequence encoding a Discosoma
CC   red fluorescent protein (DsRed) variant having a reduced propensity to
CC   oligomerize. The protein variant comprises one or more amino acid
CC   substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC   sequence, where the substitutions result in reduced propensity of the
CC   DsRed variant to form tetramers and where the variant displays detectable
CC   fluorescence of at least one red wavelength. The composition and methods
CC   are useful in producing red fluorescent proteins having reduced
CC   propensity for oligomerization, especially tetramerization. The protein
CC   may be used in molecular biology and in other scientific applications,
CC   such as in immunoassays or hybridization assays, or in tracking the
CC   movement of proteins in cells. This sequence corresponds to the DsRed
CC   variant mRFP1.1 coding sequence. The protein contains a Q66M and a T147S
CC   mutation.
XX
SQ   Sequence 678 BP; 148 A; 227 C; 208 G; 95 T; 0 U; 0 Other;

DR   WPI; 2003-845265/78.
DR   P-PSDB; ADL46281.
```

```
Alignment Scores:
Pred. No.:                2.45e-115
Score:                   1012.00
Percent Similarity:        88.69%
Best Local Similarity:     85.97%
Query Match:                  11

                    Length:    Matches:
                       678        190
                                  Mismatches:    25
                                  Indels:         0
                                  Gaps:           0
```

```
US-10-006-922A-12 (1-225) x ADL46282 (1-678)

DB:

Qy        1 MetArgSerLysLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
            | ||  ||  ||||      :: ||| ||  |||::|| | ||
Db        1 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy       21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy       41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ACCCAGACACGGCCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180
```

```
Qy       61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            ||||   || |  :: |||  | ||| |   |||||| || || ||||| || ||| |||
Db      181 CTGTCCCCTCAGTTCAGTACGGGTCCAAGGTCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy       81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            |||   || || | :| ||||| |||| ||| |||  |||  | || | |:|  ||  ||
Db      241 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy      101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            ||| || ||| ||  |||| ||||| ||||| || |  ||| ||  |:| ||||| ||
Db      301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC 360

Qy      121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            ||  ||||| ||  ||||| ||||| || ||| ||| || |||||| |||||||| |||
Db      361 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy      141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            |||||  |  || |  | |:  |||| ||  | |||||| || ||| ||||||| |||||
Db      421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy      161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            || |||||| || ||| |||| || || || |||||| ||| || || ||||| ||  ::
Db      481 ATCAAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGACCACC 540

Qy      181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            ||||| ||||||  || |||  ||| || || | ||  | ||| |:||| || || ||||
Db      541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy      201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
            || |:  || || || || |  || ||  ||| || |  | ||| || || || || |||
Db      601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC 660

Qy      221 His 221
            |||
Db      661 CAC 663
```

```
Search completed: January 12, 2006, 12:55:37
Job time : 528 secs
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

Run on: January 12, 2006, 12:26:48 ; Search time 3746 Seconds
                                           (without alignments)
                                           2810.223 Million cell updates/sec

Title:
Perfect score: 1214
Sequence: US-10-006-922A-12
                 1 MRSSKNVIKEFMRFKVRMEG..........EDYIIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop   6.0 , Fgapext 7.0
                 Delop   6.0 , Delext   7.0

Searched: 41078325 seqs, 2339541228 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 69%
                 Maximum Match 100%
                 Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat_10012006_162504_19128/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=p2nminscr.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=500
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=69 -ALIGN=500 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_1_8010 @runat_10012006_162504_19128 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                 1: gb_est1:*
                 2: gb_est2:*
                 3: gb_est3:*
                 4: gb_htc:*
                 5: gb_est4:*
                 6: gb_est5:*
                 7: gb_est6:*
                 8: gb_est7:*
                 9: gb_gss1:*
                10: gb_gss2:*
                11: gb_gss3:*

SUMMARIES

Result   Query
No.   Score Match Length DB ID   Description
---- ----- ------------ -- --   -----------

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

No matches found

Search completed: January 12, 2006, 14:59:07
Job time : 3747 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 12, 2006, 12:40:13 ; Search time 179 Seconds
(without alignments)
2234.365 Million cell updates/sec

Title: US-10-006-922A-12
Perfect Score: 1214
Sequence: 1 MRSRKNVIKEFMRFKVRMEG..........EDYTIVBQYERTEGRHHLFL 225

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 0.5
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 69%
Maximum Match 100%
Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_Bpool_p/US10006922/runat_10012006_162505_19142/app_query.fasta_1.391
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2nminscr.rnl -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cd1 -LIST=500 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=69 -ALIGN=500 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_1_290 @runat_10012006_162505_19142 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents NA:*
1: /cgn2_6/ptcdata/1/1/ina/1_COMB.seq:*
2: /cgn2_6/ptcdata/1/1/ina/5_COMB.seq:*
3: /cgn2_6/ptcdata/1/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptcdata/1/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptcdata/1/1/ina/H_COMB.seq:*
6: /cgn2_6/ptcdata/1/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptcdata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptcdata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptcdata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1214	100.0	678	3	US-09-459-956-6	Sequence 6, Appli
2	1214	100.0	859	3	US-09-866-538-11	Sequence 11, Appl
3	1214	100.0	859	3	US-09-865-291-11	Sequence 11, Appl
4	1210	99.7	723	3	US-10-152-296-1	Sequence 1, Appl
5	1210	99.7	6984	3	US-10-001-189-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

Alignment Scores:
Pred. No.: 3.24e-160   Length: 678
Score: 1214.00   Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 3   Gaps: 0

US-10-006-922A-12 (1-225) x US-09-459-956-6 (1-678)

Qy   1 MetArgSerArgLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTCAGTATGGAAGCAAAGTATATGTTAAGCACCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTCCTGAAGGATTTAAATGGGAAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTTGAACTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

```
Qy  181  TyrMetAlaLysGlyGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGGAAGGCCTGGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGlnLysTyrArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACCAACGAAGACTATACAATCGTTCAGAAGTATCGAACCGAGGGACGC 660

Qy  221  HisHisLeuPheLeu 225
         |||||||||||||
Db  661  CACCATCTGTTCCTT 675
```

RESULT 2
US-09-866-538-11
; Sequence 11, Application US/09866538
; Patent No. 6852849
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-866-538-11

Alignment Scores:
Pred. No.:
Score: 4.72e-160   Length: 859
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 100.00%   Indels: 0
                       3   Gaps: 0

US-10-006-922A-12 (1-225) x US-09-866-538-11 (1-859)

```
Qy    1  MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54  ATGAGGTCTTCCAAGAATGTTATCATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 113

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAGGGCGAAGGAAGACCCATACGGAAGGC 173

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTCGCTTGGGATATT 233

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294  GACTATAAAAAGCTGTCATTTCCTGAAGGGTTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGCTTGCAGGATGGCTGTTTCATCTAC 413

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473
```

```
Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474  ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 533

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181  TyrMetAlaLysGlyGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594  TACATGGCAAAGGAAGGACCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGlnLysTyrArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654  ATAACAAGCCACCAACGAAGACTATACAATCGTTCAGAAGTATCGAACCGAGGGACGC 713

Qy  221  HisHisLeuPheLeu 225
         |||||||||||||
Db  714  CACCATCTGTTCCTT 728
```

RESULT 3
US-09-865-291-11
; Sequence 11, Application US/09865291
; Patent No. 6900304
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: ZHANG, Jin
; APPLICANT: TING, Alice
; TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-865-291-11

Alignment Scores:
Pred. No.:
Score: 4.72e-160   Length: 859
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 100.00%   Indels: 0
                       3   Gaps: 0

US-10-006-922A-12 (1-225) x US-09-865-291-11 (1-859)

```
Qy    1  MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54  ATGAGGTCTTCCAAGAATGTTATCATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 113

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAGGGCGAAGGAAGACCCATACGGAAGGC 173

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTCGCTTGGGATATT 233

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294  GACTATAAAAAGCTGTCATTTCCTGAAGGGTTTAAATGGGAAAGGGTCATGAACTTTGAA 353
```

```
Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354  GACGGTGGCGTCGTTACTGTGACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  413

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  414  AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  473

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  474  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG  533

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  534  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  593

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  594  TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  653

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  654  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  713

Qy  221  HisHisLeuPheLeu  225
Db  714  CACCATCTGTTCCTT  728
```

RESULT 4
US-10-152-296-1
; Sequence 1, Application US/10152296
; Patent No. 6723537
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-0001100US
; CURRENT APPLICATION NUMBER: US/10/152,296
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian codon-optimized variant (DsRED) of Discosoma sp.
; FEATURE:
; OTHER INFORMATION: "red" red fluorescent protein (RFP)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(723)
; OTHER INFORMATION: DsRED
US-10-152-296-1

Alignment Scores:

Pred. No.:		
Score:	1210.00	
Percent Similarity:	100.00%	
Best Local Similarity:	99.56%	
Query Match:	3	

1.31e-159	Length:	723
	Matches:	224
	Conservative:	1
	Mismatches:	0
	Indels:	0
	Gaps:	0

US-10-006-922A-12 (1-225) x US-10-152-296-1 (1-723)

```
Qy   1  MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly  20
        ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  63

Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly  40
```

```
Db   64  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  123

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  124  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  183

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  184  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  243

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  244  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  303

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  304  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  363

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  364  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC  423

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  424  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  483

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  484  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGTATC  543

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  544  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  603

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  604  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  663

Qy  221  HisHisLeuPheLeu  225
Db  664  CACCACCTGTTCCTG  678
```

RESULT 5
US-10-001-189-45
; Sequence 45, Application US/10001189
; Patent No. 6962810
; GENERAL INFORMATION:
; APPLICANT: FRASER JR., MALCOLM J.
; APPLICANT: LI, XU
; APPLICANT: BEAM, TERESA
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
; TITLE OF INVENTION: MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
; TITLE OF INVENTION: VECTOR PIGGYBAC
; FILE REFERENCE: 835910-92098
; CURRENT APPLICATION NUMBER: US/10/001,189
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,984
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/244,677
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 45
; LENGTH: 6984
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: pBXP3-DsRed-orf sequence
US-10-001-189-45

Alignment Scores:

```
Pred. No.:                    4.87e-158     Length:         6984
Score:                        1210.00       Matches:        224
Percent Similarity:           100.00%       Conservative:   1
Best Local Similarity:        99.56%        Mismatches:     0
Query Match:                  99.67%        Indels:         0
DB:                           3             Gaps:           0

US-10-006-922A-12 (1-225) x US-10-001-189-45 (1-6984)

Qy      1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2974 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCCATGGAGGGC 3033

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3034 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3093

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3094 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3153

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3154 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3213

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3214 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3273

Qy    101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3274 GACGGCGGCGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3333

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3334 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTNATGCAGAAGAAGACC 3393

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3394 ATGGGCTGGGAGGGCCTCACCGACCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3453

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3454 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3513

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3514 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 3573

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3574 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 3633

Qy    221 HisHisLeuPheLeu 225
          |||||||||||||||
Db   3634 CACCACCTGTTCCTG 3648
```

```
                        GenCore version 5.1.6
                      Copyright (c) 1993 - 2006  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:        January 12, 2006, 12:46:57 ; Search time 792 Seconds
                                            (without alignments)
                                            2349.255 Million cell updates/sec

Title:
Perfect score: 1214
Sequence:      US-10-006-922A-12
               1 MRSSKNVIKEFMRFKVRMEG.........EDYIVEQYERTEGRHHLFL 225

Scoring table: BLOSUM62
               Xgapop  10.0 ,  Xgapext  0.5
               Ygapop  10.0 ,  Ygapext  0.5
               Fgapop   6.0 ,  Fgapext  7.0
               Delop    6.0 ,  Delext   7.0

Searched:      9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:     95

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 69%
                 Maximum Match 100%
                 Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10006922/runat 10012006 162505 19171/app_query.fasta_1.391
-DB=Published Applications NA_Main -QFMT=fastap -SUFFIX=p2nminscr.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=500 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=69 -ALIGN=500 -MODE=LOCAL -OUTFMT=pto -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_11549 @runat 10012006 162505 19171 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     Published Applications NA_Main:*
               1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
               2:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
               3:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
               4:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
               5:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
               6:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
               7:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
               8:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
               9:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
              10:  /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1214	100.0	678	3	US-09-967-772-6	Sequence 6, Appli
2	1214	100.0	678	5	US-10-006-922-11	Sequence 11, Appl
3	1214	100.0	678	7	US-10-081-864-7	Sequence 7, Appli
4	1214	100.0	678	5	US-10-121-258-2	Sequence 2, Appli
5	1214	100.0	678	6	US-10-315-920-1	Sequence 1, Appl
6	1214	100.0	678	3	US-10-132-067-3	Sequence 3, Appli
7	1214	100.0	678	6	US-10-335-517-6	Sequence 6, Appli

No.	Score	%	Length	DB	ID	Description
8	1214	100.0	678	6	US-10-334-288-6	Sequence 6, Appli
9	1214	100.0	678	7	US-10-311-030-5	Sequence 5, Appli
10	1214	100.0	678	9	US-10-656-029-21	Sequence 21, Appl
11	1214	100.0	678	9	US-10-505-486-27	Sequence 27, Appl
12	1214	100.0	678	9	US-10-844-064A-1	Sequence 1, Appli
13	1214	100.0	678	9	US-10-931-304-2	Sequence 2, Appli
14	1214	100.0	859	3	US-09-999-745-66	Sequence 66, Appli
15	1214	100.0	859	9	US-09-866-538-11	Sequence 11, Appl
16	1214	100.0	859	9	US-09-794-308-11	Sequence 11, Appl
17	1214	100.0	859	3	US-09-865-291-11	Sequence 11, Appl
18	1214	100.0	678	9	US-10-433-640-12	Sequence 12, Appl
19	1214	100.0	859	9	US-10-885-988-11	Sequence 11, Appl
20	1214	100.0	859	8	US-10-857-622-11	Sequence 11, Appl
21	1214	100.0	3311	3	US-10-006-922-35	Sequence 35, Appl
22	1210	99.7	681	9	US-10-121-258-23	Sequence 23, Appli
23	1210	99.7	681	7	US-10-311-030-8	Sequence 8, Appl
24	1210	99.7	681	9	US-10-931-304-3	Sequence 3, Appl
25	1210	99.7	681	9	US-10-931-304-23	Sequence 23, Appl
26	1210	99.7	681	9	US-10-311-030-11	Sequence 11, Appl
27	1210	99.7	681	7	US-10-311-030-12	Sequence 12, Appli
28	1210	99.7	713	7	US-10-152-296-1	Sequence 1, Appli
29	1210	99.7	713	7	US-10-152-296-1	Sequence 1, Appli
30	1210	99.7	723	7	US-10-739-856-1	Sequence 1, Appli
31	1210	99.7	723	7	US-10-214-932-51	Sequence 51, Appl
32	1210	99.7	1638	5	US-10-214-932-75	Sequence 75, Appl
33	1210	99.7	1647	9	US-10-161-403-29	Sequence 29, Appl
34	1210	99.7	4692	5	US-10-433-640-16	Sequence 16, Appl
35	1210	99.7	4692	6	US-11-006-076-29	Sequence 29, Appl
36	1210	99.7	4692	10	US-11-006-076-29	Sequence 29, Appl
37	1210	99.7	5436	7	US-10-169-050-46	Sequence 46, Appl
38	1210	99.7	6984	8	US-10-189-045	Sequence 45, Appl
39	1210	99.7	7910	8	US-10-169-050-20	Sequence 20, Appl
40	1210	99.7	9320	8	US-10-471-065-20	Sequence 20, Appli
41	1210	99.7	9658	8	US-10-609-019-4	Sequence 4, Appli
42	1210	99.7	9678	8	US-10-609-019-3	Sequence 3, Appli
43	1210	99.7	10263	8	US-10-609-019-2	Sequence 2, Appli
44	1207	99.4	678	6	US-10-609-019-3	Sequence 3, Appli
45	1203	98.7	7508	7	US-10-342-828-4	Sequence 4, Appli
46	1198	98.7	681	9	US-10-006-922-37	Sequence 37, Appl
47	1196	98.5	666	7	US-10-332-733-22	Sequence 22, Appl
48	1196	98.5	711	7	US-10-332-733-12	Sequence 12, Appli
49	1196	98.5	711	7	US-10-314-936-3	Sequence 3, Appli
50	1196	98.5	711	7	US-10-314-936-3	Sequence 3, Appli
51	1196	98.5	711	10	US-11-021-014-3	Sequence 3, Appli
52	1191	98.1	678	6	US-10-006-922-36	Sequence 36, Appl
53	1191	98.1	678	6	US-10-081-864-14	Sequence 14, Appl
54	1191	98.1	678	6	US-10-315-920-15	Sequence 15, Appl
55	1186	97.7	675	5	US-10-006-922-38	Sequence 38, Appl
56	1186	97.7	675	6	US-10-081-864-13	Sequence 13, Appl
57	1186	97.7	747	8	US-10-765-857-7	Sequence 7, Appl
58	1186	97.7	1050	9	US-10-894-949-9	Sequence 9, Appli
59	1186	97.7	4200	9	US-10-894-949-8	Sequence 8, Appli
60	1186	97.7	4300	7	US-10-742-828-5	Sequence 5, Appl
61	1186	97.7	7495	9	US-10-742-828-5	Sequence 5, Appl
62	1186	97.7	9731	9	US-10-510-363-5	Sequence 5, Appl
63	1186	97.7	9782	9	US-10-510-363-6	Sequence 6, Appl
64	1184	97.5	898	5	US-10-081-864-14	Sequence 14, Appl
65	1160	95.6	678	5	US-10-121-258-15	Sequence 5, Appl
66	1160	95.6	678	9	US-10-006-922-43	Sequence 43, Appl
67	1160	95.6	704	9	US-10-931-304-82	Sequence 82, Appl
68	1132	93.2	678	7	US-10-423-688A-40	Sequence 40, Appl
69	1121	92.3	681	9	US-10-121-258-7	Sequence 7, Appli
70	1121	92.3	681	10	US-10-931-304-7	Sequence 7, Appli
71	1121	92.3	1809	10	US-11-052-001-5	Sequence 5, Appli
72	1121	92.3	1809	9	US-10-006-922-43	Sequence 43, Appl
73	1119	92.2	678	5	US-10-081-864-11	Sequence 11, Appl
74	1119	92.2	705	9	US-10-931-304-107	Sequence 107, Appl
75	1095	90.2	1431	9	US-10-006-922-17	Sequence 17, Appl
76	1095	90.2	876	6	US-10-931-304-39	Sequence 39, Appl
77	1085.5	89.4	876	6	US-11-161-403-39	Sequence 39, Appl
78	1085.5	89.4	876	10	US-11-006-076-39	Sequence 39, Appl
79	1021	84.1	675	7	US-10-724-178-15	Sequence 15, Appl
80	1021	84.1	678	5	US-10-121-258-9	Sequence 9, Appli

No.	Score	%	Len		Accession	Sequence
81	1021	84.1	678	9	US-10-931-304-9	Sequence 9, Appli
82	1014	83.5	678	9	US-10-931-304-109	Sequence 109, App
83	1012	83.4	678	9	US-10-931-304-80	Sequence 80, App1
84	1007	82.9	678	9	US-10-931-304-105	Sequence 105, App
85	1002	82.5	678	9	US-10-931-304-101	Sequence 101, App
86	994	81.9	678	9	US-10-931-304-86	Sequence 86, App1
87	992	81.7	678	9	US-10-931-304-90	Sequence 90, App1
88	989	81.5	678	9	US-10-931-304-103	Sequence 103, App1
89	988	81.4	711	9	US-10-931-304-84	Sequence 84, App1
90	987	81.3	711	9	US-10-931-304-93	Sequence 93, App1
91	986	81.2	708	9	US-10-931-304-95	Sequence 95, App1
92	973	80.1	711	9	US-10-931-304-97	Sequence 97, App1
93	973	80.1	711	9	US-10-931-304-99	Sequence 99, App1
94	960	79.1	711	9	US-10-931-304-88	Sequence 88, App1
95	848	69.9	549	7	US-10-724-178-1040	Sequence 1040, Ap

RESULT 1

US-09-967-772-6
Sequence 6, Application US/09967772
Patent No. US20020164577A1
GENERAL INFORMATION:
APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: GONZALEZ, Jesus
TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
FILE REFERENCE: REGEN1290-5
CURRENT APPLICATION NUMBER: US/09/967,772
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US 09/459,956
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 08/765,860
PRIOR FILING DATE: 1996-12-19
PRIOR APPLICATION NUMBER: PCT/ US96/09652
PRIOR FILING DATE: 1996-06-06
PRIOR APPLICATION NUMBER: US 08/481,977
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin version 3.0
SEQ ID NO 6
LENGTH: 678
TYPE: DNA
ORGANISM: Discosoma sp "red"
US-09-967-772-6

ALIGNMENTS

US-10-006-922A-12 (1-225) x US-09-967-772-6 (1-678)

Alignment Scores:
Pred. No.:
Score: 6.83e-143   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 3   Indels: 0
       Gaps: 0

```
Qy   1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly   20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA   60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly      40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGC    120

Qy  41 HisAsnThrValLysGluLeuLysValSerGlyGlyProLeuProPheAlaTrpAspIle      60
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT    180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro      80
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA    240
```

RESULT 2

US-10-006-922-11
Sequence 11, Application US/10006922
Publication No. US20020197676A1
GENERAL INFORMATION:
APPLICANT: Lukyanov, Sergey A
APPLICANT: Fradkov, Arcady F.
APPLICANT: Labas, Yulii A.
APPLICANT: Matz, Mikhail V.
APPLICANT: Terskikh, Alexey
TITLE OF INVENTION: No. US20020197676A1e1 Chromophores/Fluorophores and
TITLE OF INVENTION: Methods for Using the Same
FILE REFERENCE: CLON-035CIP
CURRENT APPLICATION NUMBER: US/10/006,922
CURRENT FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 09/120,330
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/457,898
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,144
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,477
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/457,556
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/444,338
PRIOR FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 11
LENGTH: 678
TYPE: DNA
ORGANISM: Discosoma species
US-10-006-922-11

Alignment Scores:
Pred. No.:
Score: 6.83e-143   Length: 678
Percent Similarity: 1214.00   Matches: 225
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
       Indels: 0

```
Qy  81 AspThrTyrLysLeuLysSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTATGAACTTTGAA    300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr     120
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC    360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr     140
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA    420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu     160
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG    480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle     180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACTTAGTTGAATTCAAAAGTATT    540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp     200
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGCTACTACTATGTTGACTCCAAACTGGAT    600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg     220
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC    660

Qy 221 HisHisLeuPheLeu  225
       |||||||||||||||
Db 661 CACCATCTGTTCCTT  675
```

```
DB:                              Gaps:        0

US-10-006-922A-12 (1-225) x US-10-006-922-11 (1-678)

Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA   60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGCGAGGGCCGATACGAAGGC    120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCTTTGCTTGGGATATT   180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTCAGTATGGGAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA   240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAAGATGGCTGTTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACTTAGTTGAATTCAAAAGTATT  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGGCGC  660

Qy  221  HisHisLeuPheLeu  225
         |||||||||||||
Db  661  CACCACCTGTTCCTT  675
```

```
; SEQ ID NO 7
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-10-081-864-7

Alignment Scores:
Pred. No.:
Score:                  1214.00      Length:          678
Percent Similarity:     100.00%      Matches:         225
Best Local Similarity:  100.00%      Conservative:    0
Query Match:            100.00%      Mismatches:      0
DB:                          5       Indels:          0
                                     Gaps:            0

US-10-006-922A-12 (1-225) x US-10-081-864-7 (1-678)

Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGCGGTCTTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC   60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  660

Qy  221  HisHisLeuPheLeu  225
         |||||||||||||
Db  661  CACCACCTGTTCCTG  675
```

; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 2
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(678)
; OTHER INFORMATION: wild-type DsRed

US-10-121-258-2

Qy   Alignment Scores:
Pred. No.:
Score: 6.83e-143
Matches: 1214.00   Matches: 225
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: Gaps: 0

US-10-006-922A-12 (1-225) x US-10-121-258-2 (1-678)

Qy   1 MetArgSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCAAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACACCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACTCAAGACTCCAGCTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCCTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCCTGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCCGTTCAGCTACCAGGCTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
        ||||||||||||||||
Db 661 CACCAATCTGTTCCTT 675

RESULT 5
US-10-315-920-1
; Sequence 1, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 1
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.

US-10-315-920-1

Qy   Alignment Scores:
Pred. No.:
Score: 6.83e-143
Matches: 1214.00   Matches: 225
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: Gaps: 0

US-10-006-922A-12 (1-225) x US-10-315-920-1 (1-678)

Qy   1 MetArgSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGCGGCGTGGTGACCCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

```
Db   481 ATCCACACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp
Qy   181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   541 TACATGGCCAAGGGACAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600
Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluValGlnThrGlyGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGGTGCAGTACGGAGGCCGC 660
Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db   661 CACCACCTGTTCCTG 675
```

```
; RESULT 6
; US-10-132-067-3
; Sequence 3, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands with Intrinsic
; TITLE OF INVENTION: Fluorescence
; FILE REFERENCE: 021362-0006000US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; OTHER INFORMATION: red fluorescent protein (dsRED)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(678)
; OTHER INFORMATION: dsRED
; US-10-132-067-3

  Alignment Scores:
  Pred. No.:
  Score:                6.83e-143       Length:           678
                        1214.00         Matches:          225
  Percent Similarity:   100.00%         Conservative:     0
  Best Local Similarity: 100.00%        Mismatches:       0
  Query Match:          6               Indels:           0
  DB:                                   Gaps:             0

US-10-006-922A-12 (1-225) x US-10-132-067-3 (1-678)

Qy    1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGTACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACAATACGGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTGTCACCACAATTTCAGTATGGAAGCAAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy   81 AspThrTyrLysLeuLysPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTCATGAACTTTGAA 300
```

```
Qy  101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCGGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATGGCCAAGGGACAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluValGlnThrGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGGTACAGACGGGACGGC 660

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  661 CACCACCTGTTCCTT 675
```

```
; RESULT 7
; US-10-335-517-6
; Sequence 6, Application US/10335517
; Publication No. US20030207248A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/10/335,517
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US 09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US/09/967,772
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 08/765,860
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT/ US96/09652
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: US 08/481,977
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
; US-10-335-517-6

  Alignment Scores:
  Pred. No.:
  Score:                6.83e-143       Length:           678
                        1214.00         Matches:          225
  Percent Similarity:   100.00%         Conservative:     0
  Best Local Similarity: 100.00%        Mismatches:       0
  Query Match:          6               Indels:           0
  DB:                                   Gaps:             0

US-10-006-922A-12 (1-225) x US-10-335-517-6 (1-678)

Qy    1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
```

RESULT 8
US-10-334-288-6
Sequence 6, Application No. US10334288
Publication No. US20040002123A1
GENERAL INFORMATION:
APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: GONZALEZ, Jesus
TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
FILE REFERENCE: REGEN1290-5
CURRENT APPLICATION NUMBER: US/10/334,288
CURRENT FILING DATE: 2002-12-31
PRIOR APPLICATION NUMBER: US/09/967,772
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US 09/459,956
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 08/765,860
PRIOR FILING DATE: 1996-12-19
PRIOR APPLICATION NUMBER: PCT/ US96/09652
PRIOR FILING DATE: 1996-06-06
PRIOR APPLICATION NUMBER: US 08/481,977
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin version 3.0
SEQ ID NO 6
LENGTH: 678
TYPE: DNA
ORGANISM: Discosoma sp "red"
US-10-334-288-6

Alignment Scores:

Pred. No.:		
Score:	6.83e-143	Length: 678
	1214.00	Matches: 225
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	6	Gaps: 0

US-10-006-922A-12 (1-225) x US-10-334-288-6 (1-678)

```
Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  121  CACAATACCGTAAAGCTTAAGGTTACCAAGGGGGGACCCTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  241  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTAATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  301  GACGGTGGCGTCGTTACCGTAACCCAGGATTCCAGTTTGCAGGATGGGCTGTTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTTGAAAAGGAGGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  481  ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCACTACCTAGTTGAATTCAAAAGTATT  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  541  TACATGGCAAAGAAGCCCTGTCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  660

Qy  221  HisHisHisLeuPheLeu  225
Db  661  CACCATCTGTTCCTT  675
```

RESULT 9
US-10-311-030-5
Sequence 5, Application US/10311030
Publication No. US20040171107A1
GENERAL INFORMATION:
APPLICANT: Nelson, David
APPLICANT: Zamiara, Elize
APPLICANT: Tsien, Roger
TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
FILE REFERENCE: 15916-032US1
CURRENT APPLICATION NUMBER: US/10/311,030
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: PCT/US01/04625
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: US 60/184,732

PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
US-10-311-030-5

Alignment Scores:
Pred. No.:
Score: 6.83e-143
Percent Similarity: 1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB:

US-10-006-922A-12 (1-225) x US-10-311-030-5 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuGluMetArgPheLysPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   61 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 180

Qy   81 HisAsnThrValLysLysLeuLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnGlnMetTyrGly 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTGCTTGGGATATT 300

Qy 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTTGCCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCATCTGTTCCTT 675

RESULT 10
US-10-656-029-21
; Sequence 21, Application US/10656029
; Publication No. US20050003367A1
; GENERAL INFORMATION:
; APPLICANT: VERTEX PHARMACEUTICALS INC.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RAPID DEVELOPMENT OF
; TITLE OF INVENTION: SCREENING ASSAYS
; FILE REFERENCE: VPI/02-143W02
; CURRENT APPLICATION NUMBER: US/10/656,029
; CURRENT FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: 60/408,297
; PRIOR FILING DATE: 2002-09-05
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patentin version 3.2
US-10-656-029-21

; SEQ ID NO 21
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(678)
; OTHER INFORMATION: fluorescent protein
US-10-656-029-21

Alignment Scores:
Pred. No.:
Score: 6.83e-143
Percent Similarity: 1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB:

US-10-006-922A-12 (1-225) x US-10-656-029-21 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuGluMetArgPheLysPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy   61 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 180

Qy   81 HisAsnThrValLysLysLeuLysGlyGlyProLeuProPheAlaTrpAspIleLeu 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTGCTTGGGATATT 180

Qy 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTTGCCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTCAGTATGTCAAGCAAGGTATATGTCAAGCACACCTCCGCAGATACCA 240

Qy   41 HisAsnThrValLysLysLeuLysGlyGlyProLeuProPheAlaTrpAspIle 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValThrLysValHisProAlaAspIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTGTCACCACAATTCAGTATGGAAGCAAGGTTTGAAATAGAAGGTGTAAAGGACACATACCA 120

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ACGGTCAATGGGCACCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 120

Qy 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTTGCCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 241 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 300

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 480

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

```
Qy  221  HisHisLeuPheLeu 225
         |||||||||||||||
Db  661  CACCATCTGTTCCTT 675

RESULT 11
US-10-505-486-27
Sequence 27, Application US/10505486
; Publication No. US20050116639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 27
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
DB: US-10-505-486-27
```

Alignment Scores:
```
Pred. No.:
Score:                 6.83e-143     Length:       678
Percent Similarity:    100.00%       Matches:      225
Best Local Similarity: 100.00%       Conservative: 0
Query Match:           100.00%       Mismatches:   0
                  9                  Indels:       0
                                     Gaps:         0
```

US-10-006-922A-12 (1-225) x US-10-505-486-27 (1-678)

```
Qy    1  MetArgSerSerLysAsnValIleLeuLysGlyPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTCAGAATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACCAAGGAGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTGAAAGTTACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy   81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGluGlu  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGCTGAAGGAGGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATTCATAAGGCTCTGAAGCTTAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAAGGACGC  660

Qy  221  HisHisLeuPheLeu 225
         |||||||||||||||
Db  661  CACCATCTGTTCCTT 675

RESULT 12
US-10-844-064A-1
Sequence 1, Application US/10844064A
; Publication No. US20050149994A1
; GENERAL INFORMATION:
; APPLICANT: Bevis, Brooke
; APPLICANT: Glick, Benjamin
; TITLE OF INVENTION: RAPIDLY MATURING FLUORESCENT PROTEINS AND METHODS FOR USING THE S[
; FILE REFERENCE: 092234-9006
; CURRENT APPLICATION NUMBER: US/10/844,064A
; CURRENT FILING DATE: 2004-05-11
; PRIOR APPLICATION NUMBER: PCT/US02/40539
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US 60/341,723
; PRIOR FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma
US-10-844-064A-1
```

Alignment Scores:
```
Pred. No.:
Score:                 6.83e-143     Length:       678
Percent Similarity:    100.00%       Matches:      225
Best Local Similarity: 100.00%       Conservative: 0
Query Match:           100.00%       Mismatches:   0
                  9                  Indels:       0
DB:                                  Gaps:         0
```

US-10-006-922A-12 (1-225) x US-10-844-064A-1 (1-678)

```
Qy    1  MetArgSerSerLysAsnValIleLeuLysGlyPheMetArgPheLysValArgMetGluGly  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTCAGAATGGAAGGA  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACCAAGGAGGC  120

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CACAATACCGTAAAGCTGAAAGTTACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  180

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA  240

Qy   81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  300

Qy  101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  360

Qy  121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA  420
```

```
Qy  141 MetGlyTyrTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTCTGTATCCTCGTGATGGCGTGCTCAAAGGAGAG 480
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACTTAGTTGAATTCAAAAGTATT 540
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnThrTyrGluArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAAGAACCGAGGGACGC 660
Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

RESULT 13
US-10-931-304-2

; Sequence 2, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/866,538
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 2
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(678)
; OTHER INFORMATION: wild-type DsRed
US-10-931-304-2

Alignment Scores:
Pred. No.:
Score: 6.83e-143
                       1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 9

Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-2 (1-678)

```
Qy   1 MetArgSerLysAsnValIleLeuLysGluMetArgPheLysValArgMetGluGlyGly 20
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60
Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCGCCCATACGAAGGC 120
Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
```

RESULT 14
US-09-999-745-66
; Sequence 66, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMITTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 66
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-999-745-66

```
Db  121 CACAATACCGTAAAGCTTAAGGTTAACCAAAGGGGGGGGACCTTTGCCATTTGCCGGATTT 180
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240
Qy   81 AspTyrLeuLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300
Qy  101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCTTAGTTACTCAAGATTCAAGCCTGCAAGATGGCTGTTTCATCTAC 360
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTTAAGTTCATTGGCGTTAACTTTCCTTCCGATGGCCCTGTAATGCAAAAGAAGACA 420
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGTTAAAAGGAGAG 480
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACTTAGTTGAATTCAAAAGTATT 540
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnThrTyrGluArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAAGAACCGAGGGACGC 660
Qy  221 HisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

Alignment Scores:
Pred. No.:
Score: 9.49e-143
                       1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3

Matches: 225
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-09-999-745-66 (1-859)

Alignment Scores:
Pred. No.:

US-09-866-538-11   9.49e-143   Length: 859

RESULT 15
US-09-866-538-11
; Sequence 11, Application US/0986538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Campbell, Robert
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
; SOFTWARE: PatentIn version 3.0
; NUMBER OF SEQ ID NOS: 29
; CURRENT FILING DATE: 2001-05-24
; CURRENT APPLICATION NUMBER: US/09/866,538
; FILE REFERENCE: REGEN1530-2
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; GENERAL INFORMATION:

Qy   1   MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54   ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCGCATGGAAGGA   113

Qy   21   ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114 ACGGTCAATGGGCACCAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGGAAGGC   173

Qy   41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT   233

Qy   61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA   293

Qy   81   AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   294 GACTACTAAAAACTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA   353

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC   413

Qy   121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr   140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA   473

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG   533

Qy   161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 ATTCATAAGGCTCTGAAGAAGGATGGTGGTCATTACCTAGTTGAATTCAAAAGTATT   593

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGCTACTACTATGTTGACTCCAAACTGGAT   653

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAGAGAACCGAGGGACGC   713

Qy   221 HisHisLeuPheLeu   225
           |||||||||||||||
Db   714 CACCATCTGTTCCTT   728

RESULT 16
US-09-794-308-11
; Sequence 11, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859

US-10-006-922A-12 (1-225) x US-09-866-538-11 (1-859)

Score: 1214.00   Matches: 225
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 3   Gaps: 0

Qy   1   MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54   ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCGCATGGAAGGA   113

Qy   21   ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114 ACGGTCAATGGGCACCAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGGAAGGC   173

Qy   41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT   233

Qy   61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA   293

Qy   81   AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   294 GACTACTAAAAACTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA   353

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC   413

Qy   121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr   140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA   473

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG   533

Qy   161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 ATTCATAAGGCTCTGAAGAAGGATGGTGGTCATTACCTAGTTGAATTCAAAAGTATT   593

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGCTACTACTATGTTGACTCCAAACTGGAT   653

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAGAGAACCGAGGGACGC   713

Qy   221 HisHisLeuPheLeu   225
           |||||||||||||||
Db   714 CACCATCTGTTCCTT   728

; SEQ ID NO 11
; LENGTH: 859

Left column:

```
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (54)..(731)
US-09-794-308-11
; DB:

Alignment Scores:
Pred. No.:
Score:                      9.49e-143
Percent Similarity:         1214.00
Best Local Similarity:      100.00%
Query Match:                100.00%
                            3

                            Length:        859
                            Matches:       225
                            Conservative:  0
                            Mismatches:    0
                            Indels:        0
                            Gaps:          0

US-10-006-922A-12 (1-225) x US-09-794-308-11 (1-859)

Qy    1 MetArgSerLysSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCAAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGAATGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGGGAAGGGCCATACGAAGGC 173

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAAGTTACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTTACACCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy  221 HisHisLeuPheLeu 225
          |||||||||||||||
Db  714 CACCATCTGTTCCTT 728
```

```
RESULT 17
US-09-865-291-11
; Sequence 11, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
;   APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
;   APPLICANT: TSIEN, Roger
;   APPLICANT: TING, Alice
```

Right column:

```
;   APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NAME/KEY: CDS
;   NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
US-09-865-291-11
; DB:

SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (54)..(731)
US-09-865-291-11

Alignment Scores:
Pred. No.:
Score:                      9.49e-143
Percent Similarity:         1214.00
Best Local Similarity:      100.00%
Query Match:                100.00%
                            3

                            Length:        859
                            Matches:       225
                            Conservative:  0
                            Mismatches:    0
                            Indels:        0
                            Gaps:          0

US-10-006-922A-12 (1-225) x US-09-865-291-11 (1-859)

Qy    1 MetArgSerLysSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCAAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGAATGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGGGAAGGGGAAGGGCCATACGAAGGC 173

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAAGTTACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTTACACCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy  221 HisHisLeuPheLeu 225
          |||||||||||||||
Db  714 CACCATCTGTTCCTT 728
```

```
RESULT 18
US-10-433-640-12
Sequence 12, Application US/10433640
Publication No. US20040115792A1
GENERAL INFORMATION:
  APPLICANT: Lichtenberg-Frate, Hella
  TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
    COMPLEX ENVIRONMENTAL CONTAMINATIONS
  FILE REFERENCE: 1487/3
  CURRENT APPLICATION NUMBER: US/10/433,640
  CURRENT FILING DATE: 2003-10-02
  PRIOR APPLICATION NUMBER: PCT/EP01/14610
  PRIOR FILING DATE: 2001-12-12
  PRIOR APPLICATION NUMBER: DE 10061872.3
  PRIOR FILING DATE: 2000-12-12
  NUMBER OF SEQ ID NOS: 24
  SOFTWARE: Patentin version 3.2
  SEQ ID NO 12
  LENGTH: 859
  TYPE: DNA
  ORGANISM: Discosoma sp.
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (54)..(731)
US-10-433-640-12

Alignment Scores:
Pred. No.:
Score:                  9.49e-143    Length:         859
Percent Similarity:     100.00%      Matches:        225
Best Local Similarity:  100.00%      Conservative:   0
Query Match:            100.00%      Mismatches:     0
DB:                     7            Indels:         0
                                     Gaps:           0

US-10-006-922A-12 (1-225) x US-10-433-640-12 (1-859)

Qy    1 MetArgSerSerLysAsnValIleLeuGlyGluPheMetArgPheLysValArgMetGluGly  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAGGAGTTCATGAGGTTTAAGGTTCGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 173

Qy   41 HisAsnThrValLysGluLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACACCTGCCGACATACCA 293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 473

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ATGGGCTGGGAAGCCAGCACTGAGCGGTTGTATCCTCGTGATGGCGTGTTGAAGAGAGAG 533

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 653

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 713

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  714 CACCATCTGTTCCTT 728
```

```
RESULT 19
US-10-885-988-11
Sequence 11, Application US/10885988
Publication No. US20040259165A1
GENERAL INFORMATION:
  APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
  APPLICANT: TSIEN, Roger
  APPLICANT: Campbell, Robert
  TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
  FILE REFERENCE: REGEN1530-2
  CURRENT APPLICATION NUMBER: US/10/885,988
  CURRENT FILING DATE: 2004-07-06
  PRIOR APPLICATION NUMBER: US/09/866,538
  PRIOR FILING DATE: 2001-05-24
  NUMBER OF SEQ ID NOS: 29
  SOFTWARE: Patentin version 3.0
  SEQ ID NO 11
  LENGTH: 859
  TYPE: DNA
  ORGANISM: Discosoma sp.
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (54)..(731)
US-10-885-988-11

Alignment Scores:
Pred. No.:
Score:                  9.49e-143    Length:         859
Percent Similarity:     100.00%      Matches:        225
Best Local Similarity:  100.00%      Conservative:   0
Query Match:            100.00%      Mismatches:     0
DB:                     8            Indels:         0
                                     Gaps:           0

US-10-006-922A-12 (1-225) x US-10-885-988-11 (1-859)

Qy    1 MetArgSerSerLysAsnValIleLeuGlyGluPheMetArgPheLysValArgMetGluGly  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ATGAGGTCTTCCAAGAATGTTATCAGGAGTTCATGAGGTTTAAGGTTCGGAAGGA 113

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC 173

Qy   41 HisAsnThrValLysGluLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACACCTGCCGACATACCA 293

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Left column

```
Db  414  AAGGTCAAGTTCATTGGCGTGAACTTCCCTTCCGATGGACCTGTTATGCAAAAAGAAGACA  473
Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  474  ATGGGCTGGGAAGCCAGCACTGAACGGTTGTATCCTCGGGATGGGGTGTTGAAAGGAGAG  533
Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  534  ATTCATAAGGCTTGAAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  593
Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  594  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  653
Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  654  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  713
Qy  221  HisHisLeuPheLeu  225
Db  714  CACCATCTGTTCCTT  728
```

RESULT 20

US-10-857-622-11

; Sequence 11, Application US/10857622
; Publication No. US20050026234A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: VIOLIN, Jonathan
; APPLICANT: NEWTON, Alexandra
; APPLICANT: TSIEN, Roger
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: Emission Ratiometric Indicators of Phosphorylation By C-Kinase
; FILE REFERENCE: 39754-0891 CP1CP2
; CURRENT APPLICATION NUMBER: US/10/857,622
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: US 09/865,291
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/396,003
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: US 08/792,553
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: US 594,575
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-10-857-622-11

Alignment Scores:
Pred. No.:

Score:	9.49e-143
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	8

Length:	859
Matches:	225
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-006-922A-12 (1-225) x US-10-857-622-11 (1-859)

```
Qy   1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly  20
Db  54  ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGATTTAAGGTTCGCATGGAAGGA  113
Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly  40
Db 114  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAAGAGGAGAGGCCATACCGAAGGC  173
```

Right column

```
Qy  41  HisAsnThrValLysLeuLysValThrHisGlyGlyProLeuProPheAlaTrpAspIle  60
Db 174  CACAATACCGTAAAGCTTAAAGGTTAACCAAAGGGCGGGGGACCTTTGCCATTTGCTTGGATATT  233
Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db 234  TTGTCACCACAATTTCAGTATGGAAGCAAAGTATATGTCAAGCACCCTGCAGATATACCA  293
Qy  81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db 294  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  353
Qy 101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db 354  GACGGTGGCGTCGTTACCGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  413
Qy 121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db 414  AAGGTCAAGTTCATTGGCGTGAACTTCCCTTCCGATGGACCTGTTATGCAAAAAGAAGACA  473
Qy 141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db 474  ATGGGCTGGGAAGCCAGCACTGAACGGTTGTATCCTCGGGATGGGGTGTTGAAAGGAGAG  533
Qy 161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db 534  ATTCATAAGGCTCTCGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  593
Qy 181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db 594  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT  653
Qy 201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db 654  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC  713
Qy 221  HisHisLeuPheLeu  225
Db 714  CACCATCTGTTCCTT  728
```

RESULT 21

US-09-797-496B-3

; Sequence 3, Application US/09797496B
; Publication No. US2003004597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3311
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: ication.
US-09-797-496B-3

Alignment Scores:
Pred. No.:

Score:	6.22e-142
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	3

Length:	3311
Matches:	225
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-006-922A-12 (1-225) x US-09-797-496B-3 (1-3311)

```
Qy   1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly  20
```

```
Db   289  ATGAGGTCTTCCAAGAATGTTATCAAGGAGAGTTCATGAGGTTTAAGGTTGTTGGAAGGA  348
          |||||
Qy    21  ThrValAsnGlyHisGluPheGluIleGluIleGlyValTyrGlyGlyGly            40
Db   349  ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACGAAGGC       408
Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db   409  CACAATACCGTAAAAGCTTAAGGTTAACCAAGGGGGGACCTTTCCCATTTGCCTTGGGATATT 468
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db   469  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTTAAGCACCCTGCCGACATACCA   528
Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db   529  GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA   588
Qy   101  AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr    120
Db   589  GACGGTGGCGTCGTTACTGTGACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC   648
Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db   649  AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA   708
Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db   709  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG   768
Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db   769  ATTCATAAGGCTCTGAAGCTAAAAGACGGTGGCCATTACCTAGTTGAATTCAAAAGTATT   828
Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaSerLeuAsp  200
Db   829  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT   888
Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyGlyArg        220
Db   889  ATAACAAGCCACAACGAAGACTATACAATCGTTGAGATAACCGAGGGACGC           948
Qy   221  HisHisLeuPheLeu                                               225
Db   949  CACCACCTGTTCCTT   963
```

```
Alignment Scores:
Pred. No.:                            Length:            681
Score:            1210.00             Matches:           224
Percent Similarity: 100.00%           Conservative:        1
Best Local Similarity: 99.56%         Mismatches:          0
Query Match:      99.67%              Indels:              0
DB:                  5                Gaps:                0

US-10-006-922A-12 (1-225) x US-10-006-922-35 (1-681)

Qy     1  MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValValArgMetGluGly  20
          ::::|
Db     4  GTGCGCTCCTCCAAGAATGTTATCAAGGAGTTCATGAGGTTCAAGGTGGTGAGGATGGAGGGC     63
Qy    21  ThrValAsnGlyHisGluPheGluIleGluIleGlyValTyrGlyGlyGly                40
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64  ACGGTGAACGGGCACGAGTTCGAGATCGAGATCGGCGTGTACGGCGGCGGC              123
Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle       60
Db   124  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC     183
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro      80
Db   184  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC     243
Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu     100
Db   244  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG     303
Qy   101  AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr        120
Db   304  GACGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC       363
Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr     140
Db   364  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC     423
Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu     160
Db   424  ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAG     483
Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle     180
Db   484  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC     543
Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaSerLeuAsp     200
Db   544  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGCCTCCCTGGAC     603
Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyGlyArg          220
Db   604  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCACCGAGGGCGGC            663
Qy   221  HisHisLeuPheLeu                                                 225
Db   664  CACCACCTGTTCCTG   678
```

; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
US-10-121-258-3

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:

US-10-006-922A-12 (1-225) x US-10-121-258-3 (1-681)

			2.2e-142
			1210.00
			100.00%
			99.56%
			99.67%

Length: 681
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly 20
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCTCCTCCAAGAACGTTCATCAAGAGTTCATGCGCTTCAAGGTGCGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaLeuPro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCCTGCCC 243

Qy   81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy 221 HisHisLeuPheLeu 225
         |||||||||||||||
Db 664 CACCACCTGTTCCTG 678

RESULT 24
US-10-121-258-23
Sequence 23, Application US/10121258
Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
US-10-121-258-23

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:

US-10-006-922A-12 (1-225) x US-10-121-258-23 (1-681)

			2.2e-142
			1210.00
			100.00%
			99.56%
			99.67%

Length: 681
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValAlaArgMetGluGly 20
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCTCCTCCAAGAACGTTCATCAAGAGTTCATGCGCTTCAAGGTGCGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaLeuPro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCCTGCCC 243

Qy   81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 25
US-10-311-030-8

```
; Sequence 8, Application US/10311030
; Publication No. US20040171107A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamiara, Elize
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(678)
US-10-311-030-8

Alignment Scores:
Pred. No.:                 2.2e-142        Length:          681
Score:                     1210.00         Matches:         224
Percent Similarity:        100.00%         Conservative:    1
Best Local Similarity:     99.56%          Mismatches:      0
Query Match:               7               Indels:          0
                                           Gaps:            0

US-10-006-922A-12 (1-225) x US-10-311-030-8 (1-681)
```

RESULT 26
US-10-931-304-3

```
; Sequence 3, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
;   OTHER INFORMATION: codon usage
US-10-931-304-3

Alignment Scores:
Pred. No.:                 2.2e-142        Length:          681
Score:                     1210.00         Matches:         224
Percent Similarity:        100.00%         Conservative:    1
Best Local Similarity:     99.56%          Mismatches:      0
Query Match:               9               Indels:          0
                                           Gaps:            0

US-10-006-922A-12 (1-225) x US-10-931-304-3 (1-681)
```

; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; US-10-931-304-23

Alignment Scores:
Pred. No.:
Score: 2.2e-142   Length: 681
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
DB: 99.67%   Indels: 0
DB: 9   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-23 (1-681)

```
Qy     1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
           ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4 GTGCGCTCTCCAAGAACGTCATCAACAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy    21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy    81 AspTyrLysLysLeuSerPheProGluGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACTACAAGAAGCTGTCCTTCCCCGAGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy   161 IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ATCCACCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy   221 HisHisLeuPheLeu 225
           |||||||||||||||
Db   664 CACCACCTGTTCCTG 678
```

RESULT 27
US-10-931-304-23
; Sequence 23, Application US/10931304
; Publication No. US2005019676841
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23

RESULT 28
US-10-311-030-11
; Sequence 11, Application US/10311030
; Publication No. US2004017110741
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamiara, Elize

```
Qy     1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
           ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4 GTGCGCTCTCCAAGAACGTCATCAACAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy    21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy    81 AspTyrLysLysLeuSerPheProGluGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACTACAAGAAGCTGTCCTTCCCCGAGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy   161 IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ATCCACCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy   221 HisHisLeuPheLeu 225
           |||||||||||||||
Db   664 CACCACCTGTTCCTG 678
```

```
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/184,732
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)...(696)
US-10-311-030-11

Alignment Scores:
Pred. No.:
Score:                    2.34e-142        Length:        713
Percent Similarity:       1210.00          Matches:       224
Best Local Similarity:    100.00%          Conservative:  1
Query Match:              99.56%           Mismatches:    0
                          99.67%           Indels:        0
                                           Gaps:          0

US-10-006-922A-12 (1-225) x US-10-311-030-11 (1-713)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22 GTGAGGAGCAGCAAGAACGTGATCAAGGAGTTCATGAGGTTCAAGGTGCGGATGGAGGGC  81

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82 ACGGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGAGGGCCCCTACGAGGGC 141

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGGCCCCCTCGCCCTTCGCCTGGGACATC 201

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 202 CTGAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 261

Qy  81 AspThrTyrLysLeuSerPheProGluGlyTyrThrTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 262 GACTACAAGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 321

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 322 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCGGCTTCATCTAC 381

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 382 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 441

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 442 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 501

Qy 161 IleHisHisAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 502 ATCCACCACGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 561

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 562 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 621

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 622 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGAGGACCGAGGGCAGG 681

Qy 221 HisHisLeuPheLeu 225
        |||||||||||||||
Db 682 CACCACCTGTTCCTG 696
```

```
RESULT 29
US-10-311-030-12/c
; Sequence 12, Application US/10311030
; Publication No. US20040171107A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamiara, Elize
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/184,732
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 713
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-311-030-12

Alignment Scores:
Pred. No.:
Score:                    2.34e-142        Length:        713
Percent Similarity:       1210.00          Matches:       224
Best Local Similarity:    100.00%          Conservative:  1
Query Match:              99.56%           Mismatches:    0
                          99.67%           Indels:        0
                                           Gaps:          0

US-10-006-922A-12 (1-225) x US-10-311-030-12 (1-713)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly  20
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 692 GTGAGGAGCAGCAAGAACGTGATCAAGGAGTTCATGAGGTTCAAGGTGCGGATGGAGGGC 633

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 632 ACGGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGAGGGCCCCTACGAGGGC 573

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 572 CACAACACCGTGAAGCTTAAGGTGACCAAGGGCGGGCCCCCTCGCCCTTCGCCTGGGACATC 513

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 512 CTGAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 453

Qy  81 AspThrTyrLysLeuSerPheProGluGlyTyrThrTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 452 GACTACAAGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 393

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGlyPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 392 GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCGGCTTCATCTAC 333

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 332 AAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACC 273

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 272 ...
```

RESULT 30

US-10-152-296-1

; Sequence 1, Application US/10152296
; Publication No. US2003007730A1
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-00010US
; CURRENT APPLICATION NUMBER: US/10/152,296
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian
; OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
; OTHER INFORMATION: "red" red fluorescent protein (RFP)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(723)
; OTHER INFORMATION: DsRED

US-10-152-296-1

Alignment Scores:
Pred. No.:
Score: 2.39e-142   Length: 723
                                             Matches: 224
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 99.56%   Mismatches: 0
Query Match: 99.67%   Indels: 0
DB: 5   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-152-296-1 (1-723)

```
Db  272 ATGGGCGTGGAGGGCCTCCACCGAGCGCGCGGCGCCTGTGCCCGGCGAG 213
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||
Db  212 ATCCACAAGGCCCTGAAGCTGAAGGACGGGGGCCACTACCTGGTGGAGTTCAAGTCCATC 153
Qy  181 TyrMetAlaLeuGlyGlyProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||
Db  152 TACATGGCCCTGGGCGGGCCCGTGCAGCTGCCCGGCTACTACTACTACGTGGACTCCAAGCTGGAC 93
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyArg 220
        |||||||||||||||||||||||||||||
Db   92 ATCACCAGCCACAACGAGGACTACACCATCGTGGAGATCACCGAGGGGCAGG 33
Qy  221 HisHisLeuPheLeu 225
        ||||||||||
Db   32 CACCACCTGTTCCTG 18
```

RESULT 31

US-10-739-656-1

; Sequence 1, Application US/10739656
; Publication No. US20040126850A1
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-00010US
; CURRENT APPLICATION NUMBER: US/10/739,656
; CURRENT FILING DATE: 2003-12-17
; PRIOR APPLICATION NUMBER: US/10/152,296
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian
; OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
; OTHER INFORMATION: "red" red fluorescent protein (RFP)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(723)
; OTHER INFORMATION: DsRED

US-10-739-656-1

Alignment Scores:
Pred. No.:
Score: 2.39e-142   Length: 723
                                             Matches: 224
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 99.56%   Mismatches: 0
Query Match: 99.67%   Indels: 0
DB: Gaps: 0

US-10-006-922A-12 (1-225) x US-10-739-656-1 (1-723)

```
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||
Db  244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303
Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||
Db  304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363
Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||
Db  364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||
Db  424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483
Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        ||||||||||||||||||||||||||||||||
Db  484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543
Qy  181 TyrMetAlaLeuGlyGlyProValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||
Db  544 TACATGGCCCTGGGCGGGCCCGTGCAGCTGCCCGGCTACTACTACTACGTGGACTCCAAGCTGGAC 603
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyArg 220
        |||||||||||||||||||||||||||||
Db  604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCACCGAGGGGCAGGCACTACAGCGACGAGGGCCGC 663
Qy  221 HisHisLeuPheLeu 225
        ||||||||||
Db  664 CACCACCTGTTCCTG 678
```

```
Qy    1 MetArgSerLysValAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
        :::||:||:|||||||||||||||||||||||||||||||
Db    4 GTtCGGCTCCTCCAAGAAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63
Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        |||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123
Qy   41 HisAlaAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||
Db  124 CACGCAAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||
Db  184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243
```

```
Qy      1  MetArgSerLysAsnValIleLeuLysGlyLeuPheMetArgPheLysValArgMetGluGly    20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4  GTGCGCTTCCTCAAGAACGTCATCAACAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC    63

Qy     21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly      40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC    123

Qy     41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle      60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC    183

Qy     61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro    80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC    243

Qy     81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu    100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG    303

Qy    101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr    120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC    363

Qy    121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr    140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC    423

Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu    160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG    483

Qy    161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle    180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC    543

Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp    200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC    603

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg    220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    604  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC    663

Qy    221  HisHisHisLeuPheLeu    225
             ||||||||||||||||
Db    664  CACCACCTGTTCCTG    678
```

Alignment Scores:
```
Pred. No.:
Score:                          7.46e-142        Length:           1638
Percent Similarity:             1210.00          Matches:           224
Best Local Similarity:          100.00%          Conservative:        1
Query Match:                    99.56%           Mismatches:          0
                                99.67%           Indels:              0
                                5                Gaps:                0
DB:
```

US-10-006-922A-12 (1-225) x US-10-214-932-51 (1-1638)

```
Qy      1  MetArgSerLysAsnValIleLeuLysGlyLeuPheMetArgPheLysValArgMetGluGly    20
             :::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4  GTGCGCTTCCTCAAGAACGTCATCAACAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC    63

Qy     21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly      40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC    123

Qy     41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle      60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC    183

Qy     61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro    80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC    243

Qy     81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu    100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG    303

Qy    101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr    120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC    363

Qy    121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr    140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC    423

Qy    141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu    160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG    483

Qy    161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle    180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC    543

Qy    181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp    200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC    603

Qy    201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg    220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    604  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC    663

Qy    221  HisHisHisLeuPheLeu    225
             ||||||||||||||||
Db    664  CACCACCTGTTCCTG    678
```

; APPLICANT: KIM, Dae Heon
; APPLICANT: LEE, Yong Jik
; TITLE OF INVENTION: SYSTEM FOR DETECTING PROTEASE
; FILE REFERENCE: AP802/US
; CURRENT APPLICATION NUMBER: US/10/214,932
; CURRENT FILING DATE: 2002-08-08
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: Patentin version 3.1
;
SEQ ID NO 75
; LENGTH: 1647
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(1647)
; OTHER INFORMATION: Gene for RFP:PS(HIV-1 protease):AtOEP7:GFP hybrid protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1644)
; OTHER INFORMATION: RFP:PS(HIV-1 protease):AtOEP7:GFP hybrid protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CDS: RFP(1..678); CDS: HIV-1 protease proteolytic site(700..720);
; OTHER INFORMATION: CDS: AtOEP7(721..909); CDS: GFP(931..1644)
US-10-214-932-75

Alignment Scores:
Pred. No.:
Score: 7.52e-142
Percent Similarity: 1210.00
Best Local Similarity: 100.00%
Query Match: 99.56%
DB: 5

Length: 1647
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-214-932-75 (1-1647)

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  244 GACTACAAGAGCCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 303

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy  161 IleHisHisLysAlaLeuLysLeuLysAspGlyGlyHisHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  484 ATCCACCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543
```

RESULT 34
US-10-161-403-29
; Sequence 29, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Blena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shallard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSEQ for Windows Version 4.0
;
SEQ ID NO 29
; LENGTH: 4692
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pDsred1-N1 plasmid from Clontech
US-10-161-403-29

Alignment Scores:
Pred. No.:
Score: 3.23e-141
Percent Similarity: 1210.00
Best Local Similarity: 100.00%
Query Match: 99.56%
DB: 6

Length: 4692
Matches: 224
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-161-403-29 (1-4692)

```
Qy    1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 741

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 801

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 861

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATC 921

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922 GACTACAAGAGCCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 981
```

Below appears partial Db/Qy alignment from RESULT (earlier, right column):

```
Qy  181 TyrMetAlaLysValProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          
Db  544 TACATGGCCAAGAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          
Db  604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy  221 HisHisLeuPheLeu 225
          
Db  664 CACCACCTGTTCCTG 678
```

```
Qy   101  AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   982  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCGCTGCAGGACGGCTGCTTCATCTAC 1041

Qy   121  LysValLysPheIleGlyValAsnPheProSerArgAspGlyProValMetGlyLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1042 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCGAGGGACGGCCCCGTAATGCAGGAAGACC 1101

Qy   141  MetGlyTyrTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1102 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1161

Qy   161  IleHisLysLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1162 ATCCACAAGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1221

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1222 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1281

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1282 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCACCGAGGGCCGC 1341

Qy   221  HisHisHisLeuPheLeu 225
          ||||||||||||||||||
Db   1342 CACCACCACTGTTCCTG 1356
```

RESULT 35
US-10-433-640-16
; Sequence 16, Application US/10433640
; Publication No. US20040115792A1
; GENERAL INFORMATION:
; APPLICANT: Lichtenberg-Frate, Hella
; TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
  COMPLEX ENVIRONMENTAL CONTAMINATIONS
; FILE REFERENCE: 1487/3
; CURRENT APPLICATION NUMBER: US/10/433,640
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/14610
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: DE 10061872.3
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 16
; LENGTH: 4692
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: sequecne of vector pDSred1-N1
US-10-433-640-16

Alignment Scores:
Pred. No.:
Score: 3.23e-141   Length: 4692
                        1210.00   Matches: 224
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 99.56%   Mismatches: 0
Query Match: 99.67%   Indels: 0
DB: 7   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-433-640-16 (1-4692)

```
Qy    1  MetArgSerLysLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         :::||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682  GTGCGCCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 741

Qy   21  ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742  ACCGTGGCCAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 801

Qy   41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802  CACACACACGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 861

Qy   61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 921

Qy   81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 981

Qy  101  AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCGCTGCAGGACGGCTGCTTCATCTAC 1041

Qy  121  LysValLysPheIleGlyValAsnPheProSerArgAspGlyProValMetGlyLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1042 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCGAGGGACGGCCCCGTAATGCAGGAAGACC 1101

Qy  141  MetGlyTyrTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1102 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1161

Qy  161  IleHisLysLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1162 ATCCACAAGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1221

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1222 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1281

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1282 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCACCGAGGGCCGC 1341

Qy  221  HisHisHisLeuPheLeu 225
         ||||||||||||||||||
Db  1342 CACCACCACTGTTCCTG 1356
```

RESULT 36
US-11-006-076-29
; Sequence 29, Application US/11006076
; Publication No. US20050181506A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shelard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 17084-022002/420B
; CURRENT APPLICATION NUMBER: US/11/006,076
; CURRENT FILING DATE: 2004-12-06
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 4692
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pDSred1-N1 plasmid from Clontech
US-11-006-076-29

Alignment Scores:
Pred. No.:
Score: 3.23e-141   Length: 4692
                   1210.00   Matches: 224

Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 99.56%   Mismatches: 0
Query Match: 99.67%   Indels: 0
DB: Gaps: 0

US-10-006-922A-12 (1-225) x US-11-006-076-29 (1-4692)

```
Qy     1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db   682 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGAGGTTCAAGGTGCGCATGGAGGGC 741

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   742 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTTACGAGGGC 801

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   802 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 861

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   862 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 921

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   922 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 981

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   982 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1041

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  1042 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 1101

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  1102 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1161

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  1162 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1221

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  1222 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1281

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuArgThrGluGlyArg 220
Db  1282 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGCGCACCGAGGGCCGC 1341

Qy   221 HisHisHisLeuPheLeu 225
Db  1342 CACCACCTGTTCCTG 1356
```

```
; RESULT 37
; US-10-169-050-46
; Sequence 46, Application US/10169050
; Publication No. US20050071891A1
; GENERAL INFORMATION:
; APPLICANT: THRESHER, RON
; APPLICANT: HINDS, LYN
; APPLICANT: HARDY, CHRIS
; APPLICANT: WHYARD, STEVE
; APPLICANT: VIGNARAJAN, SOMA
; APPLICANT: GREWE, PETER MARTIN
; APPLICANT: PATIL, JAWAHAR
; TITLE OF INVENTION: REPRESSIBLE STERILITY OF ANIMALS
; FILE REFERENCE: 4050.001500
; CURRENT APPLICATION NUMBER: US/10/169,050
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: PCT/AU00/0
; PRIOR FILING DATE: 2000-12-22
```

```
; PRIOR APPLICATION NUMBER: PQ4884
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 46
; LENGTH: 5436
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pSPM6
; US-10-169-050-46
```

Alignment Scores:
Pred. No.:
Score: 1210.00
Percent Similarity: 100.00%
Best Local Similarity: 99.56%
Query Match: 99.67%
DB:

US-10-006-922A-12 (1-225) x US-10-169-050-46 (1-5436)

```
Qy     1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db  1426 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGAGGTTCAAGGTGCGCATGGAGGGC 1485

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  1486 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTTACGAGGGC 1545

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  1546 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1605

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  1606 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1665

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  1666 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1725

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  1726 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1785

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  1786 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 1845

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  1846 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1905

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  1906 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1965

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  1966 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2025

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuArgThrGluGlyArg 220
Db  2026 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGCGCACCGAGGGCCGC 2085

Qy   221 HisHisHisLeuPheLeu 225
Db  2086 CACCACCTGTTCCTG 2100
```

```
; RESULT 38
; US-10-001-189-45
; Sequence 45, Application US/10001189
```

Publication No. US20020173634A1
```
;  GENERAL INFORMATION:
;  APPLICANT:  FRASER JR., MALCOLM J.
;  APPLICANT:  LI, XU
;  APPLICANT:  BEAM, TERESA
;  TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
;  TITLE OF INVENTION:  MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
;  TITLE OF INVENTION:  VECTOR PIGGYBAC
;  FILE REFERENCE:  835910-92098
;  CURRENT APPLICATION NUMBER:  US/10/001,189
;  CURRENT FILING DATE:  2001-10-30
;  PRIOR APPLICATION NUMBER:  60/244,984
;  PRIOR FILING DATE:  2000-11-01
;  PRIOR APPLICATION NUMBER:  60/244,677
;  PRIOR FILING DATE:  2000-10-31
;  NUMBER OF SEQ ID NOS:  70
;  SOFTWARE:  Patentin Ver. 2.1
;  SEQ ID NO 45
;  LENGTH:  6984
;  TYPE:  DNA
;  ORGANISM:  Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION:  Description of Artificial Sequence:
;  OTHER INFORMATION:  pBXP3-Dsred-orf sequence
US-10-001-189-45

Alignment Scores:
Pred. No.:                5.62e-141      Length:         6984
Score:                    1210.00        Matches:        224
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    99.56%         Mismatches:     0
Query Match:              99.67%         Indels:         0
                          5              Gaps:           0
                                         DB:

US-10-006-922A-12 (1-225) x US-10-001-189-45 (1-6984)

Qy     1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2974 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3033

Qy    21 ThrValAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3034 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3093

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3094 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3153

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3154 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3213

Qy    81 AspThrTyrLysLeuLysPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3214 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3273

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3274 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3333

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3334 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 3393

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3394 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3453

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3454 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3513

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3514 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 3573

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3574 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 3633

Qy   221 HisHisLeuPheLeu 225
          |||||||||||||||
Db  3634 CACCACCTGTTCCTG 3648
```

```
RESULT 39
US-10-169-050-20/c
Sequence 20, Application US/10169050
Publication No. US2005007189191A1
;  GENERAL INFORMATION:
;  APPLICANT:  THRESHER, RON
;  APPLICANT:  HINDS, LYN
;  APPLICANT:  HARDY, CHRIS
;  APPLICANT:  WHYARD, STEVE
;  APPLICANT:  VIGNARAJAN, SOMA
;  APPLICANT:  GREWE, PETER MARTIN
;  APPLICANT:  PATIL, JAWAHAR
;  TITLE OF INVENTION:  REPRESSIBLE STERILITY OF ANIMALS
;  FILE REFERENCE:  4050.001500
;  CURRENT APPLICATION NUMBER:  US/10/169,050
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER:  PCT/AU00/0
;  PRIOR FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER:  PQ4884
;  PRIOR FILING DATE:  1999-12-24
;  NUMBER OF SEQ ID NOS:  63
;  SOFTWARE:  Patentin version 3.2
;  SEQ ID NO 20
;  LENGTH:  7910
;  TYPE:  DNA
;  ORGANISM:  Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION:  pBIT(dHSP)-RFP-oHoxDS/BH
US-10-169-050-20

Alignment Scores:
Pred. No.:                6.69e-141      Length:         7910
Score:                    1210.00        Matches:        224
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    99.56%         Mismatches:     0
Query Match:              99.67%         Indels:         0
                          9              Gaps:           0
                                         DB:

US-10-006-922A-12 (1-225) x US-10-169-050-20 (1-7910)

Qy     1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
          :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2705 GTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 2646

Qy    21 ThrValAsnGlyHisGluPheIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2645 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2586

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2585 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2526

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2525 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2466

Qy    81 AspThrTyrLysLeuLysPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2465 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 2406

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 40
US-10-471-065-20

Sequence 20, Application US/10471065
Publication No. US20040197855A1
GENERAL INFORMATION:
APPLICANT: Prof. Dr. Wiesmueller, Lisa
TITLE OF INVENTION: Test system for the determination of genotoxicities
FILE REFERENCE: P59532
CURRENT APPLICATION NUMBER: US/10/471,065
CURRENT FILING DATE: 2003-09-04
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 20
LENGTH: 9320
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Description of the artificial sequence: Plasmid
FEATURE:
OTHER INFORMATION: p5-Puro-CMV-(N'-EGFP)-CMV-Red-(EGFP-EJ)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1592)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3374)..(3392)
OTHER INFORMATION: Retroviral vector p5NM
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5527)..(9320)
OTHER INFORMATION: Retroviral vector p5NM
FEATURE:
NAME/KEY: gene
LOCATION: (1617)..(2216)
OTHER INFORMATION: Puromycin resistance gene from pRetroon (Clontech,
FEATURE:
NAME/KEY: promoter
LOCATION: (2267)..(2848)
OTHER INFORMATION: CMV promoter from pEGPP-N1 (Clontech, Palo Alto,
FEATURE:
NAME/KEY: gene
LOCATION: (2906)..(3348)
OTHER INFORMATION: N'-EGFP, derived from EGFP from pEGPP-N1
FEATURE:
NAME/KEY: gene
LOCATION: (3411)..(3992)
OTHER INFORMATION: CMV promoter from pdsRed1-N1 (Clontech, Palo Alto,
CA, USA)
FEATURE:
NAME/KEY: gene
LOCATION: (4038)..(4718)
OTHER INFORMATION: Red from pdsRed1-N1 (Clontech, Palo Alto, CA, USA)
FEATURE:
NAME/KEY: gene
LOCATION: (4766)..(5508)
OTHER INFORMATION: EGFP-EJ, derived from EGFP from pEGPP-N1 (Clontech,
Palo Alto, CA, USA)
US-10-471-065-20
```

```
Alignment Scores:
Pred. No.:
Score:                      1210.00
Percent Similarity:         100.00%
Best Local Similarity:      99.56%
Query Match:                99.67%
DB:                         8

                            8.4e-141

Length:        9320
Matches:       224
Conservative:
Mismatches:    1
Indels:        0
Gaps:          0
```

```
US-10-006-922a-12 (1-225) x US-10-471-065-20 (1-9320)

Qy    1 MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ::::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4041 GTGCGCTCCTCCAAGAAGAACGTCATCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 4100

Qy   21 ThrValAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4101 ACCGTGAACGGCCACGAGTTCGAGATCATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 4160

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleIle 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4161 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 4220

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValValAlaValLysHisProAlaAspIlePro 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4221 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGGTAGCGGTGAAGCACCCCGCCGACATCCCC 4280

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4281 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 4340

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4341 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 4400

Qy  121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4401 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 4460

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4461 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGGGTGCTGCTGAAGGGCGAG 4520

Qy  161 IleGluHisGlyAlaLeuLeuLysAspAspGlyHisTyrLeuValGluGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4521 ATCGAGCACGGCGCCCTGAAGGACGACGGCCACTACCTGGTGGAGTTCAAGTCCATC 4580

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4581 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 4640

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4641 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGC 4700

Qy  221 HisHisHisLeuPheLeu 225
         ||||||||||||||||
Db 4701 CACCACCTGTTCCTG 4715
```

```
Db 2405 GACGGGCGCGTGCGTGACCGTGACCCAGGACTCCTCCCTGCAGGAGGACGGGCTGCTTCATCTAC 2346
Qy  121 LysValLysPheIleGlyTyrValAlaAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 2345 AAGGTGAAGTTCATCGGCGTGAACTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACC 2286
         FEATURE:
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLeuTyrGlyGlu 160
Db 2285 ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGGGTGCTGCTGTACGGGGAG 2226
Qy  161 IleGluHisGlyAlaLeuLeuLysAspAspGlyHisTyrLeuValGluGluPheLysSerIle 180
Db 2225 ATCCACAAGGGCCCTGAAGGACGGGGCCACTACCTGGTGGAGTTCAAGTCCATC 2166
Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 2165 TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGGTACTACTACGTGGACTCCAAGCTGGAC 2106
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 2105 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGC 2046
Qy  221 HisHisHisLeuPheLeu 225
Db 2045 CACCACCTGTTCCTG 2031
```

RESULT 41
US-10-609-019-4

; Sequence 4, Application US/10609019
; Publication No. US20040197910A1
; GENERAL INFORMATION:
; APPLICANT: Cooper, Richard K.
; APPLICANT: Cadd, Gary G.
; APPLICANT: Fioretti, William C.
; APPLICANT: DeBoer, Kenneth F.
; TITLE OF INVENTION: Gene Regulation in Transgenic Animals Using a Transposon-Based
; TITLE OF INVENTION: Vector
; FILE REFERENCE: 51687-0101 (51687-287015)
; CURRENT APPLICATION NUMBER: US/10/609,019
; CURRENT FILING DATE: 2003-06-26
; PRIOR APPLICATION NUMBER: US 60/392,415
; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US 60/441,447
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,392
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,377
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,502
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,405
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,381
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 4
; LENGTH: 9658
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-609-019-4

Alignment Scores:
Pred. No.:
Score: 8.83e-141   Length: 9658
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
DB: 99.67%   Indels: 0
                         8   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-609-019-4 (1-9658)

```
Qy     1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5099 GTGCGCTTCCTCCAAGAAGCTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 5158

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  5159 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC 5218

Qy    41 HisAsnThrValLysGlyValThrLysGlyProLeuProPheAlaTrpAspIle 60
Db  5219 CACAACACCGTGAAGGTGACCAAGGGCGGGCCCCCTCCCTTCGCCTGGGACATC 5278

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  5279 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5338

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  5339 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 5398

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  5399 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 5458

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  5459 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 5518

Qy   141 MetGlyTrpGluAlaSerThrGluArgGluTyrProArgAspGlyValLeuLysGlyGlu 160
Db  5519 ATGGGCTGGGAGGCCTCCACCGAGCGCGAGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 5578

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  5579 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 5638

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  5639 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 5698

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  5699 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 5758

Qy   221 HisHisLeuPheLeu 225
Db  5759 CACCACCTGTTCCTG 5773
```

RESULT 42
US-10-609-019-3

; Sequence 3, Application US/10609019
; Publication No. US20040197910A1
; GENERAL INFORMATION:
; APPLICANT: Cooper, Richard K.
; APPLICANT: Cadd, Gary G.
; APPLICANT: Fioretti, William C.
; APPLICANT: DeBoer, Kenneth F.
; TITLE OF INVENTION: Gene Regulation in Transgenic Animals Using a Transposon-Based
; TITLE OF INVENTION: Vector
; FILE REFERENCE: 51687-0101 (51687-287015)
; CURRENT APPLICATION NUMBER: US/10/609,019
; CURRENT FILING DATE: 2003-06-26
; PRIOR APPLICATION NUMBER: US 60/392,415
; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US 60/441,447
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,392
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,377
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,502
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,405
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,381
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 3
; LENGTH: 9678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-609-019-3

Alignment Scores:
Pred. No.:
Score: 8.85e-141   Length: 9678
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
DB: 99.67%   Indels: 0
                         8   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-609-019-3 (1-9678)

```
Qy     1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           :::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
```

```
Db   5119 GTGCGGCTCCTCCAAGAACGTCATCAACGAGAGTTCATGCCCTTCAAGGTGCGCATGGAGGGC 5178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     21 ThrValAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40

Db   5179 ACCGTGAACGGCCACCAGTTCAAGATCGAGAGCGGCGAGGGCGAGGGCCGCTACGAGGGC 5238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     41 HisAsnThrValLysGlyLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

Db   5239 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db   5299 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

Db   5359 GACTACAAGAAGCTGTCTTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 5418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Db   5419 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 5478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db   5479 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 5538
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    141 MetGlyTyrProGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db   5539 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 5598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db   5599 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 5658
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 TyrMetAlaLeuLysLeuLysAspGlyGlyHisTyrLeuTyrTyrValAspSerSerLeuAsp 200

Db   5659 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACTACTACGTGGACTCCAGCCTGGAC 5718
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGluArgThrGluGlyArg 220

Db   5719 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGACGGAGGGCCGC 5778
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    221 HisHisHisLeuPheLeu 225

Db   5779 CACCACCTGTTCCTG 5793
          |||||||||||||||
```

```
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 2
;   LENGTH: 10263
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
; US-10-609-019-2
```

RESULT 43
US-10-609-019-2
; Sequence 2, Application US/10609019
; Publication No. US2004019791001A1
; GENERAL INFORMATION:
; APPLICANT: Cooper, Richard K.
; APPLICANT: Cadd, Gary G.
; APPLICANT: Fioretti, William C.
; APPLICANT: DeBoer, Kenneth F.
; TITLE OF INVENTION: Gene Regulation in Transgenic Animals Using a Transposon-Based
; TITLE OF INVENTION: Vector
; FILE REFERENCE: 51687-0101 (51687-287015)
; CURRENT APPLICATION NUMBER: US/10/609,019
; CURRENT FILING DATE: 2003-06-26
; PRIOR APPLICATION NUMBER: US 60/392,415
; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US 60/441,392
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,377
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,502
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,405
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,447
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,381

```
US-10-006-922A-12 (1-225) x US-10-609-019-2 (1-10263)

Alignment Scores:
Pred. No.:
Score:                        9.61e-141     Length:          10263
                              1210.00       Matches:           224
Percent Similarity:          100.00%        Conservative:        1
Best Local Similarity:        99.56%        Mismatches:          0
Query Match:                  99.67%        Indels:              0
DB:                           8             Gaps:                0
```

```
Qy      1 MetArgSerSerLysAsnValIleIleLysGlyHisPheMetArgPheLysLysArgMetGluGlyGly 20
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5725 GTGCGGCTCCTCCAAGAACGTCATCAACGAGAGTTCATGCCCTTCAAGGTGCGCATGGAGGGC 5784

Qy     21 ThrValAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5785 ACCGTGAACGGCCACGAGTTCGAGATCGAGAGCGGCGAGGGCGAGGGCCGCTACGAGGGC 5844

Qy     41 HisAsnThrValLysGlyLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5845 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5904

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5905 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5964

Qy     81 AspThrTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5965 GACTACAAGAAGCTGTCTTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 6024

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6025 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 6084

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6085 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 6144

Qy    141 MetGlyTyrProGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6145 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 6204

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6205 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 6264

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGluArgThrGluGlyArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6325 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGACGGAGGGCCGC 6384

Qy    221 HisHisHisLeuPheLeu 225
          |||||||||||||||||
Db   6385 CACCACCTGTTCCTG 6399
```

RESULT 44
US-10-315-920-3
; Sequence 3, Application US/10315920
; Publication No. US20030175809A1

; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
; US-10-315-920-3

Alignment Scores:
Pred. No.:
Score: 5.22e-142
Percent Similarity: 1207.00
Best Local Similarity: 99.56%
Query Match: 99.11%
DB: 6
                               Gaps: 0

US-10-006-922A-12 (1-225) x US-10-315-920-3 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly     20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGCGCTCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC     60

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly     40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC    120

Qy    41  HisAsnThrValLeuLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle     60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC    180

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro     80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC    240

Qy    81  AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu    100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GACTACTACAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG    300

Qy   101  AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr    120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GACGGCGGCGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC    360

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr    140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC    420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu    160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ATGGGCTGGGAGGCCTCCACCGAGCGGCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG    480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle    180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC    540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp    200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC    600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg    220
```

```
Qy   161 IleHisLysAlaLeuLeuLysGluLysGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2603 ATCCACAAGGCCCTGAAGCTGAAGGAGAAGGGCGGGCACTACCTGGTGGAGTTCAAGTCCATC 2662

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||| :::||||||||||||||||||||||||
Db  2663 TACATGGCCAAGAAGCCCGTGCAGCTCCCGGGCTACTACTACGTGGACAGCAAGCTGGAC 2722

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2723 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGGCGC 2782

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db  2783 CACCACCTGTTCCTG 2797
```

RESULT 46

US-10-006-922-37

; Sequence 37, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1e1 Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 37
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Discosoma species
; SOFTWARE: FastSEQ for Windows Version 4.0

US-10-006-922-37

Alignment Scores:
Pred. No.:
Score: 1198.00
Percent Similarity: 99.11%
Percent Local Similarity: 98.22%
Best Local Similarity: 98.68%
Query Match: 5

US-10-006-922A-12 (1-225) x US-10-006-922-37 (1-681)

Length: 681
Matches: 221
Conservative: 2
Mismatches: 2
Indels: 0
Gaps: 0

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 123

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 183
```

```
Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 243

Qy    81 AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACACCAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 303

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 483

Qy   161 IleHisLysAlaLeuLysLeuLysGlyGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ACCCACAAGGCCCTGAAGCTGAAGGGCGGGCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||| :::||||||||||||||||||||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 603

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGGCGC 663

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db   664 CACCACCTGTTCCTG 678
```

RESULT 47

US-10-332-733-22

; Sequence 22, Application US/10332733
; Publication No. US20040106565A1
; GENERAL INFORMATION:
; APPLICANT: Margarete Odenthal and Diana Jung
; TITLE OF INVENTION: Gene Expression, Genome Alteration And Reporter Expression
; TITLE OF INVENTION: In Myofibroblasts And Myofibroblast-like Cells
; FILE REFERENCE: 1472/68806
; CURRENT APPLICATION NUMBER: US/10/332,733
; CURRENT FILING DATE: 2003-06-10
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence: Red
; OTHER INFORMATION: Fluorescent Protein

US-10-332-733-22

Alignment Scores:
Pred. No.:
Score: 1196.00
Percent Similarity: 100.00%
Percent Local Similarity: 100.00%
Best Local Similarity: 98.52%
Query Match: 7

US-10-006-922A-12 (1-225) x US-10-332-733-22 (1-666)

Length: 666
Matches: 221
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```
Qy     5 LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly 24
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AAGAAGTTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGAACCGGTCAATGG 60

Qy    25 HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal 44
Db    25
```

Alignment Scores:

Pred. No.: 1.37e-140   Length: 711
Score: 1196.00   Matches: 221
Percent Similarity: 99.11%   Conservative: 2

RESULT 48
US-10-314-936-1
; Sequence 1, Application US/10314936
; Publication No. US20040110225A1
; GENERAL INFORMATION:
; APPLICANT: Gibbs, Patrick D.L.
; APPLICANT: Carter, Robert W.
; APPLICANT: Schmale, Michael C.
; TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
; FILE REFERENCE: 638.004
; CURRENT APPLICATION NUMBER: US/10/314,936
; CURRENT FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: mutant red fluorescent protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(711)
; OTHER INFORMATION:
US-10-314-936-1

Best Local Similarity: 98.22%
Query Match: 98.52%
DB:

     Mismatches: 2
     Indels: 0
     Gaps: 0

US-10-006-922A-12 (1-225) x US-10-314-936-1 (1-711)

RESULT 49
US-10-314-936-3
; Sequence 3, Application US/10314936
; Publication No. US20040110225A1
; GENERAL INFORMATION:
; APPLICANT: Gibbs, Patrick D.L.
; APPLICANT: Carter, Robert W.
; APPLICANT: Schmale, Michael C.
; TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
; FILE REFERENCE: 638.004
; CURRENT APPLICATION NUMBER: US/10/314,936
; CURRENT FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:

Left column

```
;   OTHER INFORMATION: mutant red fluorescent protein
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (1)..(711)
;   OTHER INFORMATION:
US-10-314-936-3

Alignment Scores:
Pred. No.:
Score:                    1.37e-140    Length:        711
Score:                    1196.00      Matches:       221
Percent Similarity:       99.11%       Conservative:  2
Best Local Similarity:    98.22%       Mismatches:    2
Query Match:              98.52%       Indels:        0
DB:                                    Gaps:          0

US-10-006-922A-12 (1-225) x US-10-314-936-3 (1-711)

Qy    1 MetArgSerLysLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1 ATGAGGTCGTTCCAAGAATGTTATCAAGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 60

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGGAGGAGAAGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTAAAGCATCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAACGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCGTTACTGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGTTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGGAGGACGC 660

Qy  221 HisHisHisLeuPheLeu 225
Db  661 CACCATCTGTTCCTT 675
```

Right column

```
;   TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
;   FILE REFERENCE: 638.004
;   CURRENT APPLICATION NUMBER: US/11/021,014
;   CURRENT FILING DATE: 2004-12-23
;   ORGANISM: Artificial
;   PRIOR APPLICATION NUMBER: US/10/314,936
;   PRIOR FILING DATE: 2002-12-09
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: Patentin version 3.1

SEQ ID NO 1
;   LENGTH: 711
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: mutant red fluorescent protein
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (1)..(711)
;   OTHER INFORMATION:
US-11-021-014-1

Alignment Scores:
Pred. No.:
Score:                    1.37e-140    Length:        711
Score:                    1196.00      Matches:       221
Percent Similarity:       99.11%       Conservative:  2
Best Local Similarity:    98.22%       Mismatches:    2
Query Match:              98.52%       Indels:        0
DB:                                    Gaps:          0

US-10-006-922A-12 (1-225) x US-11-021-014-1 (1-711)

Qy    1 MetArgSerLysLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db    1 ATGAGGTCGTTCCAAGAATGTTATCAAGAGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 60

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGGAGGAGAAGGCCATACGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db  181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTAAAGCATCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAACGGGTCATGAACTTTGAA 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  301 GACGGTGGCGTCGTTACTGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCTGATGGACCTGTTATGCAAAAGAAGACA 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db  421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAGGGAGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  481 ATTCATAAGGCTCTGAAGTTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db  541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db  601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGGAGGACGC 660
```

RESULT 50
US-11-021-014-1
; Sequence 1, Application US/11021014
; Publication No. US20050100954A1
; GENERAL INFORMATION:
;   APPLICANT: Gibbs, Patrick D.L.
;   APPLICANT: Carter, Robert W.
;   APPLICANT: Schmale, Michael C.

Qy 221 HisHisHisLeuPheLeu 225
        |||||||||||||||||
Db 661 CACCATCTGTTCCTT 675

RESULT 51
US-11-021-014-3
; Sequence 3, Application US/11021014
; Publication No. US20050100954A1
; GENERAL INFORMATION:
; APPLICANT: Gibbs, Patrick D.L.
; APPLICANT: Carter, Robert W.
; APPLICANT: Schnale, Michael C.
; TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
; FILE REFERENCE: 638.004
; CURRENT APPLICATION NUMBER: US/11/021,014
; CURRENT FILING DATE: 2004-12-23
; PRIOR APPLICATION NUMBER: US/10/314,936
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: mutant red fluorescent protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(711)
; OTHER INFORMATION:
US-11-021-014-3

Alignment Scores:
Pred. No.:
Score: 1.37e-140   Length: 711
Score: 1196.00   Matches: 221
Percent Similarity: 99.11%   Conservative: 2
Best Local Similarity: 98.22%   Mismatches: 2
Query Match: 98.52%   Indels: 0
DB: 10   Gaps: 0

US-10-006-922A-12 (1-225) x US-11-021-014-3 (1-711)

Qy   1 MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTTGTTCCAAGAATAGTTATCAAGGAGGTTCATGAGGTTTAAGGTTCGTATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGGCACCAGAGTTTGAAATAGAAGGCGAAGGAGAAGGCCRTACGAAGGAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAACAAAGGTATATGTTCAAGCATCCTGCCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTCCAGCCTCCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCCGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 52
US-10-006-922-36
; Sequence 36, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US2002019767676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-36

Alignment Scores:
Pred. No.:
Score: 5.47e-140   Length: 678
Score: 1191.00   Matches: 220
Percent Similarity: 98.67%   Conservative: 2
Best Local Similarity: 97.78%   Mismatches: 3
Query Match: 98.11%   Indels: 0
DB: 5   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-006-922-36 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly 20
        ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db   1 ATGGCGCTCCTCCGCAGGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 421 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||:|||||||||||||||||:|||||||||||||||||:|||||||||||||
Db 481 ATTCATAAGGCTCTGAAGTTGAAAGACCGGTGGTCATTTACCTAGTTGAATTCAAAACTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||:|||:|||||||||||:||||||||||||||||||||||||||:||||||||
Db 541 TACATGGCAAAGAAGCCCGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||:||:||:||||||||||||||||||||||||:|||||
Db 601 ATAACAAGCCACAACAAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAAGGACGC 660

Qy 221 HisHisHisLeuPheLeu 225
        |||||||||||||||
Db 661 CACCATCTGTTCCTT 675

Alignment Scores:
Pred. No.:
Score: 5.47e-140
Percent Similarity: 98.67%
Best Local Similarity: 97.78%
Query Match: 98.11%

US-10-081-864-14
SEQ ID NO 14
LENGTH: 678
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: non-aggregating mutant
SOFTWARE: FastSEQ for Windows Version 4.0
NUMBER OF SEQ ID NOS: 30
PRIOR APPLICATION NUMBER: 60/270,983
PRIOR APPLICATION FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 10/006,922
PRIOR APPLICATION FILING DATE: 2001-12-04
CURRENT FILING DATE: 2002-06-19
CURRENT APPLICATION NUMBER: US/10/081,864
FILE REFERENCE: CLON-067
TITLE OF INVENTION: Methods for Using the Same
TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
APPLICANT: Fradkov, Arcady
APPLICANT: Savitsky, Alexander
APPLICANT: Yanushevich, Yuriy
APPLICANT: Lukyanov, Konstantin
APPLICANT: Lukyanov, Sergey
GENERAL INFORMATION:
Publication No. US20030022287A1
Sequence 14, Application US/10081864

RESULT 53

US-10-081-864-14

Db 121 CACACACCGTGAAGTTGAAGGTGACCAAGGGCGGGCGCCCCTCGCCTTGCCCTGGGACATC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db 181 CTGTCCCCCAGTTCCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCACATCCCC 240
Qy   81 AspTyrLysLysLeuSerPheProGlyLysPheLysTrpGluArgValMetAsnPheGlu 100

Qy 241 GACTACAAGAAGCTGTCCTTCCCGGAGGGGTTCAAGTGGGGAGCGGGTGATGAACTTCGAG 300
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Db 301 GACGGCGGCGTGGCGACCGTGACCCAGGACTCCTCCCTGCAGGACGGGCTGCTTCATCTAC 360
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTGATGCAGAAGAAGACC 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db 421 ATGGGCTGGGAGGCCCTCCACCGAGCGGCTGTACCCCCGCGACGGGGTCCTGAAGGGCGAG 480
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Qy 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCGGGCTACTACTACGTGGACAGCAAGCTGGAC 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnThrGluGlyArg 220

Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGCAGACGGAGGGCCGC 660
Qy 221 HisHisHisLeuPheLeu 225

Db 661 CACCACCTGTTCCTG 675

DB: US-10-006-922A-12 (1-225) x US-10-081-864-14 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyGly 20
        ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCCTCCTCCAAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
Db   61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAlaAsnThrValLysGluLysLysValLysLysGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACGCCAACACCGTGAAGTTGAAGGTGAAGGTGGGGCCCCTGCCCTTCGCGTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCAGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAAGCTGTCCTTCCCGGAGGGGTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGGCCCGTGATGCAGAAGAAGACC 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAGGCCCTCCACCGAGCGGCTGTACCCCCGCGACGGGGTGCTGAAGGGCGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCCAAGAAGCCCGTGCAGCTGCCGGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlnThrGluGlyArg 220
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGCAGACGGAGGGCCGC 660

Qy 221 HisHisHisLeuPheLeu 225
Db 661 CACCACCTGTTCCTG 675

RESULT 54

US-10-315-920-5

Sequence 5, Application US/10315920
Publication No. US20030175809A1
GENERAL INFORMATION:
APPLICANT: Fradkov, Arcady Fedorovich
APPLICANT: Terskikh, Alexey
TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
TITLE OF INVENTION: FOR THEIR USE
FILE REFERENCE: CLON-077CIP
CURRENT APPLICATION NUMBER: US/10/315,920
CURRENT FILING DATE: 2002-12-09
PRIOR APPLICATION NUMBER: 60/211,607
PRIOR APPLICATION FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: PCT/US01/19097
PRIOR APPLICATION FILING DATE: 2001-06-13
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 678
TYPE: DNA

Length: 678
Matches: 220
Conservative: 2
Mismatches: 3
Indels: 0

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-5

Alignment Scores:
Pred. No.:
Score: 5.47e-140   Length: 678
Percent Similarity: 1191.00   Matches: 220
Best Local Similarity: 98.67%   Conservative: 2
Query Match: 97.73%   Mismatches: 3
DB: 98.11%   Indels: 0
                        6   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-315-920-5 (1-678)

Qy   1   MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly   20
             |||||||||||||||::|||||||||||||||||||||||||||||||||||||||||||
Db   1   ATGGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTNCGCATGGAGGGC   60

Qy   21   ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   120

Qy   41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   180

Qy   61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   240

Qy   81   AspGlyTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC   360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC   420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG   480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC   540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC   600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuValArgThrGluGlyArg   220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGGTGCGCACCGAGGGCCGC   660

Qy   221 HisHisHisLeuPheLeu   225
             ||||||||||||||||||
Db   661 CACCACCTGTTCCTG   675

ORGANISM: Discosoma species
US-10-006-922-38

; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-38

Alignment Scores:
Pred. No.:
Score: 2.32e-139   Length: 675
Percent Similarity: 1186.00   Matches: 219
Best Local Similarity: 98.22%   Conservative: 2
Query Match: 97.33%   Mismatches: 4
DB: 97.69%   Indels: 0
                        5   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-006-922-38 (1-675)

Qy   1   MetArgSerSerLysAsnValIleLeuGluPheMetArgPheLysValArgMetGluGly   20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   ATGGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTNCGCATGGAGGGC   60

Qy   21   ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   120

Qy   41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   180

Qy   61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   240

Qy   81   AspGlyTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC   360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC   420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG   480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC   540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC   600

; RESULT 55
US-10-006-922-38
; Sequence 38, Application No. US20020197676A1
; Publication No. US200201976761A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey

RESULT 56

US-10-081-864-13

```
; Sequence 13, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
;  APPLICANT: Lukyanov, Sergey
;  APPLICANT: Lukyanov, Konstantin
;  APPLICANT: Yanushevich, Yuriy
;  APPLICANT: Savitsky, Alexandr
;  APPLICANT: Fradkov, Arcady
;  TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
;  TITLE OF INVENTION: Methods for Using the Same
;  FILE REFERENCE: CLON-067
;  CURRENT APPLICATION NUMBER: US/10/081,864
;  CURRENT FILING DATE: 2002-06-19
;  PRIOR APPLICATION NUMBER: 10/006,922
;  PRIOR FILING DATE: 2001-12-04
;  PRIOR APPLICATION NUMBER: 60/270,983
;  PRIOR FILING DATE: 2001-02-21
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;  LENGTH: 675
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: non-aggregating mutant
US-10-081-864-13

  Alignment Scores:
  Pred. No.:
  Score:                    2.32e-139        Length:          675
  Percent Similarity:       1186.00          Matches:         219
  Percent Similarity:       98.22%           Conservative:      2
  Best Local Similarity:    97.33%           Mismatches:        4
  Query Match:              97.69%           Indels:            0
                                             Gaps:              0
```

US-10-006-922A-12 (1-225) x US-10-081-864-13 (1-675)

```
Qy         1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
             ||||||||||||||||||||||||||||||||||||||| :|:|||||||||||||||||
Db         1 ATGGCCTCCTCCAGAAGCATCACCAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC    60

Qy        21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy        41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy        61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy        81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTCATGAACTTCGAG  300

Qy       101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360
```

RESULT 57

US-10-785-862-10

```
; Sequence 10, Application US/10785862
; Publication No. US20040261149A1
; GENERAL INFORMATION:
;  APPLICANT: Donald Danforth Plant Science Center
;  APPLICANT: Fauquet, Claude M.
;  APPLICANT: Padmanabhan, Chellappan
;  APPLICANT: Ramachandran, Vanitharani
;  TITLE OF INVENTION: siRNA-mediated inhibition of gene expression in plant cells
;  FILE REFERENCE: C35621/ 104850
;  CURRENT APPLICATION NUMBER: US/10/785,862
;  CURRENT FILING DATE: 2004-02-24
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;  LENGTH: 747
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: the dsRed coding sequence
US-10-785-862-10

  Alignment Scores:
  Pred. No.:
  Score:                    2.68e-139        Length:          747
  Percent Similarity:       1186.00          Matches:         219
  Percent Similarity:       98.22%           Conservative:      2
  Best Local Similarity:    97.33%           Mismatches:        4
  Query Match:              97.69%           Indels:            0
                                             Gaps:              0
```

US-10-006-922A-12 (1-225) x US-10-785-862-10 (1-747)

```
Qy         1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
             ||||||||||||||||||||||||||||||||||||||| :|:|||||||||||||||||
Db         1 ATGGCCTCCTCCAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC    60

Qy        21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy        41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy        61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240
```

```
Qy    81 AspTyrLysGlyGlyLeuSerPheProGlyPheLysTrpGluArgValMetAsnPheGlu 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTACAAGAAGGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACGGCGGCGTGGGCGACCGTGACCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy   181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACATGGCCAAGGGCCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy   221 HisHisLeuPheLeu 225
            |||||||||||||||
Db   661 CACCACCTGTTCCTG 675
```

RESULT 58
US-10-060-857-7
; Sequence 7, Application US/10060857
; Publication No. US20020132318A1
; GENERAL INFORMATION:
; APPLICANT: AntiCancer, Inc.
; APPLICANT: Zhao, Ming
; APPLICANT: Jiang, Ping
; APPLICANT: Xu, Mingxu
; APPLICANT: Yang, Meng
; TITLE OF INVENTION: IMPROVED FLUORESCENT PROTEINS
; FILE REFERENCE: 31276-20032.00
; CURRENT APPLICATION NUMBER: US/10/060,857
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 60/264,932
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1050
; TYPE: DNA
; ORGANISM: Coral
US-10-060-857-7

Alignment Scores:
Pred. No.:
Score: 1186.00   Length: 1050
Percent Similarity: 98.22%   Matches: 219
Best Local Similarity: 97.33%   Conservative: 2
Query Match: 97.69%   Mismatches: 4
DB: 4.3e-139   Indels: 0
                                           Gaps: 0

US-10-006-922A-12 (1-225) x US-10-060-857-7 (1-1050)

```
Qy     1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
            |||||||||||||||||||||||||||||||||||||||||||||::||||||||||
Db     1 ATGGGCTCCTCCGAGAACGTCATCACCGAGTTCATGGGCTTCAAGGTGCGCATGGAGGGC 60

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   349 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 408
```

```
Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   409 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 468

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   469 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 528

Qy    81 AspTyrLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   529 GACTACAAGGGCCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 588

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   589 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 648

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   649 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 708

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   709 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 768

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   769 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 828

Qy   181 TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   829 TACATGGCCAAGGGCCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 888

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   889 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 948

Qy   221 HisHisLeuPheLeu 225
            |||||||||||||||
Db   949 CACCACCTGTTCCTG 963
```

RESULT 59
US-10-894-949-9
; Sequence 9, Application US/10894949
; Publication No. US20050130919A1
; GENERAL INFORMATION:
; APPLICANT: Xia, Xuguang
; APPLICANT: Xu, Zuoshang
; TITLE OF INVENTION: REGULATABLE PROMOTERS FOR SYNTHESIS OF SMALL HAIRPIN RNA
; FILE REFERENCE: UMY-072
; CURRENT APPLICATION NUMBER: US/10/894,949
; CURRENT FILING DATE: 2004-07-19
; PRIOR APPLICATION NUMBER: 60/488,510
; PRIOR FILING DATE: 2003-07-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 4200
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: expression construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 528, 2516
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 528..2516
US-10-894-949-9

Alignment Scores:
Pred. No.:
Score: 1186.00   Length: 4200
Percent Similarity: 98.22%   Matches: 219
Best Local Similarity: 97.33%   Conservative: 2
                         2.96e-138   Mismatches: 4

```
Query Match:        97.69%      Indels:     0
DB:                     9       Gaps:       0

US-10-006-922A-12 (1-225) x US-10-894-949-9 (1-4200)

Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||::|||||||||||||||||||||||||||||||||||||||||||
Db 3267 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 3326

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3327 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 3386

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3387 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 3446

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3447 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3506

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3507 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3566

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3567 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3626

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3627 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 3686

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3687 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 3746

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db 3747 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3806

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db 3807 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 3866

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlyGluArgLeuGluGlyGlu 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3867 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGGGCGAGCGCCTGGAGGGCGAG 3926

Qy  221 HisHisLeuPheLeu 225
          |||||||||||||||
Db 3927 CACCACCTGTTCCTG 3941
```

```
RESULT 60
US-10-894-949-8
; Sequence 8, Application US/10894949
; Publication No. US20050130919A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Zuoshang
; APPLICANT: Xia, Xugang
; TITLE OF INVENTION: REGULATABLE PROMOTERS FOR SYNTHESIS OF SMALL HAIRPIN RNA
; FILE REFERENCE: UMY-072
; CURRENT APPLICATION NUMBER: US/10/894,949
; CURRENT FILING DATE: 2004-07-19
; PRIOR APPLICATION NUMBER: 60/488,510
; PRIOR FILING DATE: 2003-07-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 4300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: expression construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 536, 2369
; OTHER INFORMATION: n = A,T,C or G
US-10-894-949-8
DB:

US-10-006-922A-12 (1-225) x US-10-894-949-8 (1-4300)

Alignment Scores:
Pred. No.:
Score:                 1186.00       Length:        4300
Percent Similarity:    98.22%        Matches:        219
Best Local Similarity: 97.33%        Conservative:     2
Query Match:           97.69%        Mismatches:       4
                                     Indels:           0
                                     Gaps:             0
```

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
          |||||||||||||||::|||||||||||||||||||||||||||||||||||||||||||
Db 1291 ATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 1350

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1351 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 1410

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1411 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 1470

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1471 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 1530

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1531 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 1590

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1591 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 1650

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1651 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 1710

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1711 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 1770

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
          ||||||||||||||||||||||||||||||||||||::||||||||||||||||||||||
Db 1771 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 1830

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
          |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db 1831 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 1890

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluLeuGlyGluArgLeuGluGlyGlu 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1891 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCTGGGCGAGCGCCTGGAGGGCGAG 1950

Qy  221 HisHisLeuPheLeu 225
          |||||||||||||||
Db 1951 CACCACCTGTTCCTG 1965
```

```
RESULT 61
US-10-742-828-5
; Sequence 5, Application US/10742828
; Publication No. US20040157294A1
; GENERAL INFORMATION:
; APPLICANT: Huynh, Gigi
; APPLICANT: Heinrich, Gerhard
; TITLE OF INVENTION: TRANSGENIC SCREEN AND METHOD FOR SCREENING MODULATORS OF BRAIN-DE
```

; FILE REFERENCE: US 1353/03 (VA)
; CURRENT APPLICATION NUMBER: US/10/742,828
; CURRENT FILING DATE: 2003-12-23
; PRIOR APPLICATION NUMBER: US 10/306,737
; PRIOR FILING DATE: 2002-11-29
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 5
; LENGTH: 7495
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: Unsure
; LOCATION:
; OTHER INFORMATION: Synthesized
US-10-742-828-5

Alignment Scores:
Pred. No.:
Score: 1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 97.69%
DB:

                6.64e-138
Length: 7495
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-742-828-5 (1-7495)

```
Qy     1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2123 ATGGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGGCTTCAAGGTGCGCATGGAGGGC 2182

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2183 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 2242

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2243 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2302

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2303 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 2362

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2363 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 2422

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2423 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 2482

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2483 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 2542

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGluGlu  160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2543 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 2602

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2603 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCTATC 2662

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         |||||||||||||||||||||||||||||||||||||| ::||||||||||||||||||||
Db  2663 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 2722

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2723 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 2782

Qy   221 HisHisLeuPheLeu  225
         |||||||||||||||
Db  2783 CACCACCTGTTCCTG 2797
```

RESULT 62
US-10-510-363-5
; Sequence 5, Application US/10510363
; Publication No. US20050158724A1
; GENERAL INFORMATION:
; APPLICANT: Fishel, Richard A.
; APPLICANT: Yoder, Kristine E.
; TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS THAT MODULATE A DNA REPAIR
; TITLE OF INVENTION: PATHWAY AND/OR RETRO-VIRAL INFECTIVITY, THE COMPOUNDS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SCHN-0033
; CURRENT APPLICATION NUMBER: US/10/510,363
; CURRENT FILING DATE: 2004-10-05
; PRIOR APPLICATION NUMBER: PCT/US03/10302
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: US 60/370,376
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 5
; LENGTH: 9731
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthesized retroviral vectors
US-10-510-363-5

Alignment Scores:
Pred. No.:
Score: 1186.00
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 97.69%
DB:

                9.55e-138
Length: 9731
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-510-363-5 (1-9731)

```
Qy     1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5496 ATGGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGGCTTCAAGGTGCGCATGGAGGGC 5555

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5556 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 5615

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5616 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 5675

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5676 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 5735

Qy    81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5736 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 5795

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5796 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 5855

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5856 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGAAGACC 5915

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGluGlu  160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5916 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 5975

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5976 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCTATC 6035
```

RESULT 63
US-10-510-363-6

; Sequence 6, Application US/10510363
; Publication No. US20050158724A1
; GENERAL INFORMATION:
; APPLICANT: Fishel, Richard A.
; APPLICANT: Yoder, Kristine E.
; TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS THAT MODULATE A DNA REPAIR
; TITLE OF INVENTION: PATHWAY AND/OR RETRO-VIRAL INFECTIVITY, THE COMPOUNDS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SCHN-0033
; CURRENT APPLICATION NUMBER: US/10/510,363
; CURRENT FILING DATE: 2004-10-05
; PRIOR APPLICATION NUMBER: 2003-04-04
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/10302
; PRIOR FILING DATE: 2003-04-05
; PRIOR APPLICATION NUMBER: US 60/370,376
; NUMBER OF SEQ ID NOS: 6
; PRIOR FILING DATE: 2002-04-05
; SOFTWARE: Patentin version 3.3
;
; SEQ ID NO 6
; LENGTH: 9782
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthesized retroviral vectors
US-10-510-363-6

Alignment Scores:
Pred. No.:
Score: 9.62e-138
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 9

Length: 9782
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-510-363-6 (1-9782)

Qy   1 MetArgSerSerLysAsnValIleLysGluProPheMetArgPheLysValArgMetGluGly 20
         ||||| :::||||||||||||:::||||||||||||||||||||||||||
Db   1 ATGGCCTTCCTCCGAGAACGTCATCACCGAGTTCATACGGCTTCAAGGTTCGGCATGGAGGGC 72

Qy   21 ThrValAsnGlyHisGlnPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||
Db   73 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 132

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         |||||||||||||||||||||||||||||||||||||||||||
Db 133 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 192

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||
Db 193 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 252

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||
Db 253 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 312

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         ||||||||||||||||||||||||||||||||||||||
Db 313 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 372

RESULT 64
US-10-006-922-45

; Sequence 45, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A.
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1e1 Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 45
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-45

Alignment Scores:
Pred. No.:
Score: 6.18e-139
Percent Similarity: 98.22%
Best Local Similarity: 97.33%
Query Match: 5

Length: 898
Matches: 219
Conservative: 2
Mismatches: 4
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-006-922-45 (1-898)

Qy   1 MetArgSerSerLysAsnValIleLysGluProPheMetArgPheLysValArgMetGluGly 20
         ||||| :::||||||||||||:::||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 152

Qy 101 ...

Alignment Scores:

```
Qy    21  ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   153  ACGGTCAAATGGGCACCAGTTTGAAATAGAAGGCGAAGGGGAAGGCCATACGAAGGC  212

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   213  CACAATACCGTAAAAGTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT  272

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   273  TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACACCCGCCGACATACCA  332

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   333  GACTATAAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA  392

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   393  GACGGTGGCGTGGTCACTGTTACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC  452

Qy   121  LysValLysLeuArgGlyThrAsnPheProSerAspGlyProValMetGlnLysLysThr  140
             ||:|||||||||||||||||||::: |||||||||||||||||||||||||||||||||
Db   453  AAGTCAAGTTCATTGGCGGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA  512

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   513  ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGGTGAAGGGAGAG  572

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   573  ATTCATAAGGCTCTGAAGCTAAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT  632

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   633  TACATGGCAAAGAAGCCTGTGCAGCTACCAGGATACTACTATGTTGACTCCAAAACTGAT  692

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyGluGlyArg  220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   693  ATAACAAGCCACAACGAAGACTATACAAATCGTTGAGCAGTTATTGAAAGAGACCGGAGGCGC  752

Qy   221  HisHisLeuPheLeu  225
             |||||||||||||||
Db   753  CACCACCTGTTCCTT  767
```

Pred. No.:   4.47e-136   Length: 678
Score: 1160.00   Matches: 216
Percent Similarity: 97.33%   Conservative: 3
Best Local Similarity: 96.00%   Mismatches: 6
Query Match: 95.55%   Indels: 0
DB: 5   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-121-258-5 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
             |||
Db     1  ATGGCCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy    21  ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyGluGlyArgProTyrGluGly  40
             |:::|||||||||:|:::|||||||||||||||||||||||||||||||||||||||||
Db    61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
             ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  ACCAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC  240

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyGluPheIleTyr  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC  360

Qy   121  LysValLysLeuArgGlyThrAsnPheProSerAspGlyProValMetGlnLysLysThr  140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT  420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC  600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluIleThrGluGlyGluGlyArg  220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGATCACCGAGGGCGAGGGCCGC  660

Qy   221  HisHisLeuPheLeu  225
             |||||||||||||||
Db   661  CACCACCTGTTCCTG  675
```

; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "T1"
US-10-931-304-5

Alignment Scores:
Pred. No.:
Score: 4.47e-136
                         1160.00
Percent Similarity: 97.33%
Best Local Similarity: 96.00%
Query Match: 95.55%
DB:

Length: 678
Matches: 216
Conservative: 3
Mismatches: 6
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-5 (1-678)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
          |||
Db    1 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  60

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
          :::||
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAspThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 420

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  421 ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db  601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy  221 HisHisLeuPheLeu  225
          |||||
Db  661 CACCACCTGTTCCTG 675
```

RESULT 67
US-10-844-064A-3
; Sequence 3, Application US/10844064A
; Publication No. US20050149994A1
; GENERAL INFORMATION:
; APPLICANT: Bevis, Brooke
; APPLICANT: Glick, Benjamin
; TITLE OF INVENTION: RAPIDLY MATURING FLUORESCENT PROTEINS AND METHODS FOR USING THE SA
; FILE REFERENCE: 092234-9006
; CURRENT APPLICATION NUMBER: US/10/844,064A
; CURRENT FILING DATE: 2004-05-11
; PRIOR APPLICATION NUMBER: PCT/US02/40539
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US 60/341,723
; PRIOR FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 3
; LENGTH: 704
; TYPE: DNA
; ORGANISM: Discosoma
US-10-844-064A-3

Alignment Scores:
Pred. No.:
Score: 4.72e-136
                         1160.00
Percent Similarity: 97.33%
Best Local Similarity: 96.00%
Query Match: 95.55%
DB:

Length: 704
Matches: 216
Conservative: 3
Mismatches: 6
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-844-064A-3 (1-704)

```
Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly  20
          |||
Db   20 ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC  79

Qy   21 ThrValAlaAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
          :::||
Db   80 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 139

Qy   41 HisAspThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  60
Db  140 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 199

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  80
Db  200 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 259

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db  260 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 319

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db  320 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 379

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db  380 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT 439

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db  440 ATGGGCTGGGAGGCCTCCACGGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 499

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db  500 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 559

Qy  181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db  560 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 619
```

```
Qy  201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  620 ATCACCTCCCACAACAGGAGTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGC 679

Qy  221 HisHisLeuPheLeu 225
        |||||||||||||||
Db  680 CACCACCTGTTCCTG 694
```

RESULT 68
US-10-423-688A-40
; Sequence 40, Application US/10423688A
; Publication No. US20040078148A1
; GENERAL INFORMATION:
; APPLICANT: Los Alamos National Laboratory
; APPLICANT: Waldo, Geoffrey S.
; TITLE OF INVENTION: Directed Evolution Methods for Improving Polypeptide Folding and
; TITLE OF INVENTION: Solubility and Superfolder Fluorescent Proteins Generated Thereb
; FILE REFERENCE: S-100,608
; CURRENT APPLICATION NUMBER: US/10/423,688A
; CURRENT FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: 10/132,067
; PRIOR FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 40
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
US-10-423-688A-40

Alignment Scores:
Pred. No.:
Score: 1132.00   Length: 678
Percent Similarity: 95.56%   Matches: 208
Best Local Similarity: 92.44%   Conservative: 7
Query Match: 93.25%   Mismatches: 10
                                        Indels: 0
                                        Gaps: 0

US-10-006-922A-12 (1-225) x US-10-423-688A-40 (1-678)

```
Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
       |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAGTCTTCCGAGGANGTTAATCAAGGAGGAGTTCATGAGGTTTAAGGTTCACATGGAGGA 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
       :::||||:||||:|||:|||||||:|||||:||||||||||||||||||||:|||:|||
Db  61 TCGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGAGGCCATACGAAGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
       |||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 ACCCAGGAACGTAAAGCTTAAGGTAACTAAGGGGGGACCTTTGCCCAATTGCTTGGGATATT 180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
       ||||||||||:||:|||||||||||||||||||||||||||:::|||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy  81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
       ||||||||:||||||||||||||||||||||:||:|||||||||||||:|:|||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
       |||||||||||:||||||:||||||||||:||:||||||||:||||||||||||||
Db 301 GACGGTGGCGTCGTCACTGTAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
       |||||||:|||||||||||:|||||||||||||||:||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCCGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
       ||||||||||:||||||||||||||||:|||||||||:|:|:|||:|:::
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAT 480
```

```
Qy 161 IleHisGlyAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
       |||||||:||:||||||||||||::::
Db 481 ATTCATGGAGGCTTCGAAGCCTGAAAGACCTGGAACGGGTGGTCATTTACCTTAGTGAATATCAAAGTCATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
       :::|||||||||||:|:|:|||||||:||||:|||:||:::|||:::|||
Db 541 TACATGGCAAAGAAGCCTGTGCAGCTACCAGGTACTACTATGTTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
       |||||||||||||||:||:|||||||||:::|:|||||:|:|||:|||
Db 601 ATAACAAGACCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAGCCGAGGGACGC 660

Qy 221 HisHisLeuPheLeu 225
       |||||||||||||||
Db 661 CACCACCTGTTCCTT 675
```

RESULT 69
US-10-121-258-7
; Sequence 7, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Sequence
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
US-10-121-258-7

Alignment Scores:
Pred. No.:
Score: 1121.00   Length: 681
Percent Similarity: 95.96%   Matches: 207
Best Local Similarity: 92.83%   Conservative: 7
Query Match: 92.34%   Mismatches: 9
                                        Indels: 0
                                        Gaps: 0

US-10-006-922A-12 (1-225) x US-10-121-258-7 (1-681)

```
Qy   3 SerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
       ||||||||:||::|:||||||||||||||||||||||:||||||||||||||:
Db  10 TCCTCCGAGGACGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGGATGGAGGGCTCCGTG 69

Qy  23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
       :||||||||||||:|||||:|||||||||||||||||||||||||||:|||||:
Db  70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCAG 129

Qy  43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
       |||||||||||||||||:||||||||||||||||||||||||||||:||:
Db 130 ACGGTCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 189

Qy  63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
       ||||:||||||||||||||||:|||||||||:|||:::|:|||||:||||
Db 190 CCCCAGTTCCAGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGACTAC 249

Qy  83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
       ||||||||||||||||:|||||||||:|||||||||||:|||:||||:|||||:
Db 250 AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 309
```

RESULT 70

```
US-10-931-304-7
; Sequence 7, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
;  APPLICANT: Tsien, Roger
;  APPLICANT: Campbell, Robert
;  APPLICANT: Baird, Geoffrey
;  TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;  TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;  FILE REFERENCE: 39754-0831CP2CP3
;  CURRENT APPLICATION NUMBER: US/10/931,304
;  CURRENT FILING DATE: 2004-08-30
;  PRIOR APPLICATION NUMBER: 10/209,208
;  PRIOR FILING DATE: 2002-07-29
;  PRIOR APPLICATION NUMBER: 10/121,258
;  PRIOR FILING DATE: 2002-04-10
;  PRIOR APPLICATION NUMBER: 09/866,538
;  PRIOR FILING DATE: 2001-05-24
;  PRIOR APPLICATION NUMBER: 09/794,308
;  PRIOR FILING DATE: 2001-02-26
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;  LENGTH: 681
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
US-10-931-304-7
```

```
Alignment Scores:
Pred. No.:
Score:                    3.77e-131      Length:        681
                          1121.00        Matches:       207
Percent Similarity:       95.96%         Conservative:  7
Best Local Similarity:    92.83%         Mismatches:    9
Query Match:              92.34%         Indels:        0
DB:                       9              Gaps:          0
```

```
US-10-006-922A-12 (1-225) x US-10-931-304-7 (1-681)

Qy   3 SerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
        ||||||| || ||| |||  || |||  ||||:::||| |||||||
Db  10 TCCTCCGAGGACGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTG 69

Qy  23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAAC 129

Qy  43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
        ||||||||||||||||| ||||||||||||||||||||||||||
Db 130 ACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 189

Qy  63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
        ||||||||||||||||||||||||||||||||||||||||||||
Db 190 CCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTAC 249

Qy  83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
        ||||||||||||||||||||||||||||||||||||||||||||
Db 250 AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 309

Qy 103 GlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIleTyrLysVal 122
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 310 GGCGTGGTGACGGTGACCCAGGACGGCTGCTTCATCTACAAGGTG 369

Qy 123 LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 370 AAGTTCCGCGGCGCCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC 429

Qy 143 TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis 162
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 430 TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC 489

Qy 163 LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet 182
        :::||||||||||||||||||||||||||||||||||||||||||
Db 490 CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATG 549

Qy 183 AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr 202
        ||||||||||||||||||||||||||||||||||||||||||||
Db 550 GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACC 609

Qy 203 SerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHis 222
        |||||||||||||||||||||:::|||||||||||||||||||||||
Db 610 TCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGCCACCAC 669

Qy 223 LeuPheLeu 225
        |||||||||
Db 670 CTGTTCCTG 678
```

RESULT 71

```
US-11-052-001-5
; Sequence 5, Application US/11052001
; Publication No. US20050233356A1
; GENERAL INFORMATION:
;  APPLICANT: JONES, JOSHUA T.
;  APPLICANT: HAHN, ANGELA TERESA
;  APPLICANT: MEYER, TOBIAS
;  TITLE OF INVENTION: LIVE-CELL BIOSENSOR POLYPEPTIDES AND
;  TITLE OF INVENTION: METHODS OF USE
;  FILE REFERENCE: STAN-385
;  CURRENT APPLICATION NUMBER: US/11/052,001
;  CURRENT FILING DATE: 2005-02-03
;  PRIOR APPLICATION NUMBER: 60/541,297
;  PRIOR FILING DATE: 2004-02-03
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;  LENGTH: 1809
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: expression vector
US-11-052-001-5
```

```
Alignment Scores:
Pred. No.:
Score:            1.47e-130     Length:   1809
                  1121.00       Matches:  207
```

```
Qy  23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAAC 129

Qy  43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
        ||||||||||||||||| ||||||||||||||||||||||||||
Db 130 ACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 189

Qy  63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
        ||||||||||||||||||||||||||||||||||||||||||||
Db 190 CCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTAC 249

Qy  83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
        ||||||||||||||||||||||||||||||||||||||||||||
Db 250 AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 309

Qy 103 GlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIleTyrLysVal 122
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 310 GGCGTGGTGACGGTGACCCAGGACGGCTGCTTCATCTACAAGGTG 369

Qy 123 LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 370 AAGTTCCGCGGCGCCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC 429

Qy 143 TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis 162
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 430 TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC 489

Qy 163 LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet 182
        :::||||||||||||||||||||||||||||||||||||||||||
Db 490 CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATG 549

Qy 183 AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr 202
        ||||||||||||||||||||||||||||||||||||||||||||
Db 550 GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACC 609

Qy 203 SerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHis 222
        |||||||||||||||||||||:::|||||||||||||||||||||||
Db 610 TCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCACCGAGGGCCGCCACCAC 669

Qy 223 LeuPheLeu 225
        |||||||||
Db 670 CTGTTCCTG 678
```

Percent Similarity: 95.96%   Conservative: 7
Best Local Similarity: 92.83%   Mismatches: 9
Query Match: 92.34%   Indels: 0
DB: 10   Gaps: 0

US-10-006-922A-12 (1-225) x US-11-052-001-5 (1-1809)

Qy   3   SerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrVal   22
       ||||||||:|||||||||||||||||||||||||||||||:|||
Db   424   TCCTCCGAGAACGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCGTCCGTG   483

Qy   23   AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn   42
       |||||||||||||||||||||||||||||||||||||||||||||
Db   484   AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCCAG   543

Qy   43   ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer   62
       |||||||||||||||||||||||||||||||||||||||||||||
Db   544   ACCGGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC   603

Qy   63   ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr   82
       |||||||||||||||||||||||||||||||||||||||||||||
Db   604   CCCCAGTTCCAGTACGGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGACTAC   663

Qy   83   LysGlyLeuSerPheProGlyGlyPheLysTrpGluArgValMetAsnPheGluAspGly   102
       |||||||||||||||||||||||||||||||||||||||||||||
Db   664   AAGAAGCTGTCCTTCCCCGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC   723

Qy   103   GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal   122
       |||||||||||||||||||||||||||||||||||||||||||||
Db   724   GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGACTGCTTCATCTACAAGGTG   783

Qy   123   LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly   142
       |||||||||||||||||||||||||||||||||||||||||||||
Db   784   AAGTTCATCGGCGACACCAACTTCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC   843

Qy   143   TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis   162
       |||||||||||||||||||||||||||||||||||||||||||||
Db   844   TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC   903

Qy   163   LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet   182
       :::||||||||||||||||||||||||||||||||||||
Db   904   CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATG   963

Qy   183   AlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr   202
       |||||||||||||||||||||||||||||||||||||||||||||
Db   964   GCCAAGAGAGCCCGTGCCAGCGTACCGCTACTACTACGTGGACAGCAAGCTGGACATCACC   1023

Qy   203   SerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHisHis   222
       |||||||||||||||||||||||||||||||||||||||||:|:|
Db   1024   TCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCACCAC   1083

Qy   223   LeuPheLeu   225
       |||||||||
Db   1084   CTGTTCCTG   1092

RESULT 72
US-10-006-922-43
; Sequence 43, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898

; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hybrid construct
US-10-006-922-43

US-10-006-922A-12 (1-225) x US-10-006-922-43 (1-678)

; Alignment Scores:
Pred. No.:   6.7e-131
Score: 1119.00
Percent Similarity: 96.89%
Best Local Similarity: 91.56%
Query Match: 92.17%
DB: 5

   Length: 678
   Matches: 206
   Conservative: 12
   Mismatches: 7
   Indels: 0
   Gaps: 0

Qy   1   MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
       ||||||||||||||||||||||||||||||||||||||||||||
Db   1   ATGAGCTCGAGCAAGAACGTGATCAAGGAGTTCATGCGGTTCAAGGTGCGGATGGAGGGC   60

Qy   21   ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
       |||:|||||||||||||||||||||||||||||||||||||||||
Db   61   ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   120

Qy   41   HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
       |:::||||||||||||||||||||||||||||||||||||||||
Db   121   CACTGCAGCGTGAAGCTCATGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   180

Qy   61   LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
       ||||||||||||||||||||||||||||||||||||||||||||
Db   181   CTCAGCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   240

Qy   81   AspTyrLysLysLeuSerPheProGlyGlyPheLysTrpGluArgValMetAsnPheGlu   100
       |||||||||||||||||||||||||||||||||||||||||||||
Db   241   GACTACAAGAAGCTCAGCTTCCCCGGCGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG   300

Qy   101   AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
       |||||||||||||||||||||||||||||||||:::|||||||||
Db   301   GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGCTGCTTCATCTAC   360

Qy   121   LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
       |||||||||||||||||||||||||||||||||||||||||||||
Db   361   GAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGCCCCGTGATGCAGAAGAAGACG   420

Qy   141   MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
       |||||||||||||||||||||||||||||||||||||||:::||
Db   421   CGGGGGCTGGGAGGCCAGCACGGAGCGGCTCTACCCCCGGGACGGCGTGCTGAAGGGCGAC   480

Qy   161   IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
       ||||||||||||||||:|:||||||||||||||||||||||||||
Db   481   ATCCACAAGGCCCTCCGGCTCAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC   540

Qy   181   TyrMetAlaLysGlyProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
       |||||||||||||||||||||||||||||||||||||||||||||
Db   541   TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACAGCAAGCTGGAC   600

Qy   201   IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
       |||||||||||||||||||||||||||||||||||||||||||:|:
Db   601   ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCCGG   660

```
Qy    221 HisHisHisLeuPheLeu 225
          |||||||||||||||||
Db    661 CACCACCTCTTCCTC 675
```

RESULT 73
US-10-081-864-11

; Sequence 11, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Pradkov, Arcady
; TITLE OF INVENTION: Methods for Using the Same
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hybrid coding sequence
US-10-081-864-11

Alignment Scores:
Pred. No.:
Score: 6.7e-131
Percent Similarity: 1119.00
Best Local Similarity: 96.89%
Query Match: 91.56%
                              92.17%
                              5

Length: 678
Matches: 206
Conservative: 12
Mismatches: 7
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-081-864-11 (1-678)

```
Qy     1  MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
          ATGAGCTGCAGCAAGAACGTGATCAAGGAGTTCATGCGGTTCAAGGTGCGGATGGGAGGGC 60
Db     1  
```

```
Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db    61  ACCGTGAACGGCCACGAGTTCGAGATCAAGGGCGAGGGCGAGGGCGCCCTACGAGGGC 120
```

```
Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   121  CACTGCAGCGTGAAGCTCATGGTGACCAAGGGCGGCGCCCCTTCGCTTCGACATC 180
```

```
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   181  CTCTCAGCCCCCCAGTTCCAGTACGGCAGCAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
```

```
Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   241  GACTACAAGAAGCTCAGCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300
```

```
Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   301  GACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTCCAAGACGGCTGCTTCATCTAC 360
```

```
Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   361  GAGGTGAAGTTCATCGGCGTGAACTTCCCCAGCGACGGGCCCCGTGATGCAGAAGAAGACC 420
```

```
Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   421  
```

```
Db   421  CGGGGCTGGGAGGCCAGCAGCGAGCGGCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAC 480
```

```
Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   481  ATCCACAATGCCCTCAAGCTGAAGGACGGCGGCCACTACCTCGTGGAGTTCAAGAGCATC 540
```

```
Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   541  TACATGGCCAAGAAGCCCGTGCAGCTCCCCGGCTACTACTACGTGGACAGCAAGCTCGAC 600
```

```
Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   601  ATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGGACGGAGGGCCG 660
```

```
Qy   221  HisHisHisLeuPheLeu 225
          |||||||||||||||||
Db   661  CACCACCTCTTCCTC 675
```

RESULT 74
US-10-931-304-82

; Sequence 82, Application US/10931304
; Publication No. US20050167681A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 705
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynuc. of Polypeptide variant "dimer2.2MMM(dimer3)(dtomato)"
US-10-931-304-82

Alignment Scores:
Pred. No.:
Score: 7.5e-128
Percent Similarity: 1095.00
Best Local Similarity: 94.57%
Query Match: 90.95%
                              90.20%
                              9

Length: 705
Matches: 201
Conservative: 8
Mismatches: 12
Indels: 0
Gaps: 0
DB:

US-10-006-922A-12 (1-225) x US-10-931-304-82 (1-705)

```
Qy     5  LysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGly 24
          :::                |||||||||||||||||||||||::::|| |||||
Db    16  GAGGGAGGTCATCAAAGAGTTCATGCGCTTCAAAGTGCGCATGGAGGGCTCCATGAACGGC 75
```

```
Qy    25  HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal 44
Db    76  CACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAACACCGCC 135
```

```
Qy    45  LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln 64
Db   136  AAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAG 195
```

```
Qy    65  PheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLys 84
          |||                                                       |||
Db   
```

Alignment Scores:

Pred. No.:		
Score:	2.03e-127	Length:	1431
Percent Similarity:	1095.00	Matches:	201
Best Local Similarity:	94.57%	Conservative:	8
Query Match:	90.95%	Mismatches:	12
DB:	90.20%	Indels:	0
	9	Gaps:	0

RESULT 75
US-10-931-304-107
; Sequence 107, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; PRIOR APPLICATION NUMBER: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "cdTomato"
US-10-931-304-107

US-10-006-922A-12 (1-225) x US-10-931-304-107 (1-1431)

```
Qy     5 LysAsnValIleLysGlyPheMetArgPheLysValArgMetGluGlyThrValAsnGly 24
              |||                  |:::                  |||
Db    16 GAGGAGGTCATCAAAAGATTCAATGGCGTTCAAGATGGCCATGGAGGGCTCCATGAACGGC 75

Qy    25 HisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrVal 44
         |||                                     |||
Db    76 CACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGACCCTACGAGGGCCACAACACCGTC 135

Qy    45 LysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGln 64
         |||                                              |||
Db   136 AAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAG 195

Qy    65 PheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLys 84
         |||                                              |||
Db   196 TTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAG 255

Qy    85 LeuSerPheProGluGlyPhelysTrpGluArgValMetAsnPheGluAspGlyGlyVal 104
         |||                                    |:::          |||
Db   256 CTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAAGACGGCGGCGTG 315

Qy   105 ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe 124
         |||                                              |:::
Db   316 GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGATG 375

Qy   125 IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu 144
         |||                                              |||
Db   376 CGCGGCGCACCAACTTCCCCCCGGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG 435

Qy   145 AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAla 164
         |||                                              |||
Db   436 GCCTCCACCGAGCGCCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAGATCCACAAGGCC 495

Qy   165 LeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys 184
         |||                                              |:::
Db   496 CTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATGGCCAAG 555

Qy   185 LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis 204
         |||                                    |:::
Db   556 AAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACCTCCCAC 615

Qy   205 AsnGluAspTyrThrIleValGluGlnTyrGluArgHisGluGlyLeuLysLeuProGlu 224
         |||                                              |:::
Db   616 AACGAGGACTACACCATCGTGGAACAGTACGAGCGGCCCGAGGGCCTGAAGCTGCCCGTTC 675

Qy   225 Leu 225
         |||
Db   676 CTG 678
```

RESULT 76
US-10-006-922-17
; Sequence 17, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1e1 Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09

US-10-006-922A-12

```
Qy   196 TTCATGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAG 255

Qy    85 LeuSerPheProGluGlyPhelysTrpGluArgValMetAsnPheGluAspGlyGlyVal 104
            |||                                 |:::          |||
Db   256 CTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAAGACGGCGGCGTG 315

Qy   105 ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe 124
         |||                                              |:::
Db   316 GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGATG 375

Qy   125 IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu 144
         |||                                              |||
Db   376 CGCGGCGCACCAACTTCCCCCCGGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG 435

Qy   105 ValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPhe 124
         |||                                              |:::
Db   316 GTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGATG 375

Qy   125 IleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGlu 144
         |||                                              |||
Db   376 CGCGGCGCACCAACTTCCCCCCGGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG 435

Qy   145 AlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAla 164
         |||                                              |||
Db   436 GCCTCCACCGAGCGCCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAGATCCACAAGGCC 495

Qy   165 LeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLys 184
         |||                                              |:::
Db   496 CTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATGGCCAAG 555

Qy   185 LysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHis 204
         |||                                    |:::
Db   556 AAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACCTCCCAC 615

Qy   205 AsnGluAspTyrThrIleValGluGlnTyrGluArgHisGluGlyLeuLysLeuProGlu 224
         |||                                              |:::
Db   616 AACGAGGACTACACCATCGTGGAACAGTACGAGCGGCCCGAGGGCCTGAAGCTGCCCGTTC 675

Qy   225 Leu 225
         |||
Db   676 CTG 678
```

Left column

```
;  PRIOR APPLICATION NUMBER: 09/457,556
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/444,338
;  PRIOR FILING DATE: 1999-11-19
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 17
;  LENGTH: 876
;  TYPE: DNA
;  ORGANISM: Discosoma species
US-10-006-922-17
```

```
Alignment Scores:
Pred. No.:
Score:                    1.62e-126        Length:        876
                          1085.50          Matches:       199
Percent Similarity:       95.58%           Conservative:   17
Best Local Similarity:    88.05%           Mismatches:      9
Query Match:              89.42%           Indels:          1
DB:                       5                Gaps:            1
```

US-10-006-922A-12 (1-225) x US-10-006-922-17 (1-876)

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
          ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   45 ATGAGTTGTTCCAAGAATGTGATCAAGGAGTTCATGAGGTTCGTATGGAAGGA 104

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
          ||||||||||||||||||||||::|||||||||||||||||||||||
Db  105 ACGGTCAATGGGCACGAGTTTGAAATAAAAGGCGAAGGTGAAGGGCCTTACGAAGGT 164

Qy   41 HisAsnThrValGlyLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          :::|||||||    |||||||||||||||||||||||||||||||
Db  165 CACTGTTCCGTAAAGCTTATGGTAACCAAGGGTGGACCTTTGCCATTTGCTTTTGATATT 224

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          |||||||||||||||||||||||||||||||||||||||||
Db  225 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTAAAGCACCCTGCCGACATACCA 284

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
          ||||||||||||||||||||||||||||||||||||||||
Db  285 GACTATAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 344

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          |||||||||||::|||||||||||||||||||||||||||||
Db  345 GACGGTGGCGTGGTTACTGTTACCCAAGATTCCAGTTTGCAGGACGGCTGTTTCATCTAC 404

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
          :::|||||||||||||||||||||||||||||||||||
Db  405 GAGGTCAAGTTCATTGGGGTGAACTTTCCTTCGTGATGGACCTGTTATGCAGAAGAAGACA 464

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
          ||||||||||||||||||||||||||||||||::|
Db  465 CGGGGCGCTGGGAAGCCAGCTACCGAGCGGTTGTATCCTCGTGATGGGGTGCTGAAAGGAGAC 524

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
          ||||||||||||||||||::|||||
Db  525 ATCCATATGGCTCTGAAGGCGGGGCCATTACCTCGTTGAATTCAAAAGTATT 584

Qy  181 TyrMetAlaLysLysProValGlnLeuProGly---ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeu  199
          ||||||||||||
Db  585 TACAAGGTAAAGACCCTTCAGTGCAGTTCCCAGGCTACTATTATGTTGACTCCAAACTG 644

Qy  200 AspIleThrSerHisAsnGluAspTyrThrIleValGluIleGluGlnTyrGluArgThrGluGly  219
          ||||||||
Db  645 GATATGACGAGCCACAACGAAGATTACACAGTCGTTGAGGAGTATGAAAAAACCCAGGGA 704

Qy  220 ArgHisHisLeuPheLeu  225
          ||||||
Db  705 CGCCACCACCGGTTCATT 722
```

Right column

```
Alignment Scores:
Pred. No.:
Score:                    1.62e-126        Length:        876
                          1085.50          Matches:       199
Percent Similarity:       95.58%           Conservative:   17
Best Local Similarity:    88.05%           Mismatches:      9
Query Match:              89.42%           Indels:          1
DB:                       6                Gaps:            1
```

US-10-006-922A-12 (1-225) x US-10-161-403-39 (1-876)

```
Qy    1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
          ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   45 ATGAGTTGTTCCAAGAATGTGATCAAGGAGTTCATGAGGTTCGTATGGAAGGA 104

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
          ||||||||||||||||||||||::|||||||||||||||||||||||
Db  105 ACGGTCAATGGGCACGAGTTTGAAATAAAAGGCGAAGGTGAAGGGCCTTACGAAGGT 164

Qy   41 HisAsnThrValGlyLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          :::|||||||    |||||||||||||||||||||||||||||||
Db  165 CACTGTTCCGTAAAGCTTATGGTAACCAAGGGTGGACCTTTGCCATTTGCTTTTGATATT 224

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          |||||||||||||||||||||||||||||||||||||||||
Db  225 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTAAAGCACCCTGCCGACATACCA 284

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
          ||||||||||||||||||||||||||||||||||||||||
Db  285 GACTATAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 344

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
          |||||||||||::|||||||||||||||||||||||||||||
Db  345 GACGGTGGCGTGGTTACTGTTACCCAAGATTCCAGTTTGCAGGACGGCTGTTTCATCTAC 404

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
          :::|||||||||||||||||||||||||||||||||||
Db  405 GAGGTCAAGTTCATTGGGGTGAACTTTCCTTCGTGATGGACCTGTTATGCAGAAGAAGACA 464

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
          ||||||||||||||||||||||||||||||||::|
Db  465 CGGGGCGCTGGGAAGCCAGCTACCGAGCGGTTGTATCCTCGTGATGGGGTGCTGAAAGGAGAC 524
```

```
RESULT 78
US-11-006-076-39
; Sequence 39, Application US/11006076
; Publication No. US20050181506A1
; GENERAL INFORMATION:
;  APPLICANT: Perkins, Edward
;  APPLICANT: Perez, Carl
;  APPLICANT: Lindenbaum, Michael
;  APPLICANT: Greene, Amy
;  APPLICANT: Leung, Josephine
;  APPLICANT: Fleming, Elena
;  APPLICANT: Stewart, Sandra
;  APPLICANT: Shellard, Joan
;  TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
;  FILE REFERENCE: 17084-022002/420B
;  CURRENT APPLICATION NUMBER: US/11/006,076
;  CURRENT FILING DATE: 2004-12-06
;  PRIOR APPLICATION NUMBER: 60/294,758
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: 60/366,891
;  PRIOR FILING DATE: 2002-03-21
;  NUMBER OF SEQ ID NOS: 129
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 39
;  LENGTH: 876
;  TYPE: DNA
;  ORGANISM: Discosoma species
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (45)...(737)
;  OTHER INFORMATION: Nucleotide sequence encoding red flourescent
;  PUBLICATION INFORMATION: protein (FP53)
US-11-006-076-39

  Alignment Scores:
  Pred. No.:
  Score:                    1.62e-126       Length:        876
                            1085.50         Matches:       199
  Percent Similarity:       95.58%          Conservative:   17
  Best Local Similarity:    88.05%          Mismatches:      9
  Query Match:              89.42%          Indels:          1
  DB:                       10              Gaps:            1

US-10-006-922A-12 (1-225) x US-11-006-076-39 (1-876)

Qy    1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyGly 20
          |  | | | | ||  |  || | |  |   | |||  | ||  | |||  ||| ||
Db   45 ATGAGTTGTTCCAAGAATGTGATCAAGGAGTTCATGAGGTTCAAGGTTCGTATGGAAGGA 104

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
          ||  |  | | |  |  |||  | |  |  |  |  |  |  ||| || |   ||  ||
Db  105 ACGGTCAATGGGCACGAGTTTGAAATAGAAGGCGAAGGTGAGGGAAGGCCCTACGAAGGT 164
```

```
Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspAspValLeuLysGlyGlu 160
Db  465 CGGGGCTGGGAAGCCAGCTCTGAGGCGTTGTATCCTCGTGATGGGGTGCTGAAAGGAGAC 524

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  525 ATCCATATGGGCTCTGAGGCGGGAGGCGGCCATTACCTCGTTGAATTCAAAAGTATT 584

Qy  181 TyrMetAlaLysLysPro---ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeu 199
Db  585 TACATGGTAAAGAAGCCTTCAGTGCAGTTGCCAGGCTACTACTATTATGTTGACTCCAAACTG 644

Qy  200 AspIleThrSerHisAsnGluAspTyr......GluLysThrGlnGly 219
Db  645 GATATGACGAGCCACAACGAAGATTACACAGTCGTTGAGCAGTATGATGAAAAAACCCAGGGA 704

Qy  220 ArgHisHisLeuPheLeu 225
          ||| ||| ||| | |::
Db  705 CGCCACCATCCGTTCATT 722
```

```
RESULT 79
US-10-724-178-15
; Sequence 15, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
;  APPLICANT: Odyssey Thera, Inc.
;  APPLICANT: Michnick, Stephen
;  APPLICANT: MacDonald, Marnie
;  APPLICANT: Lamerdin, Jane
;  TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
;  COMPLEMENTATION ASSAYS
;  FILE REFERENCE: ODY007
;  CURRENT APPLICATION NUMBER: US/10/724,178
;  CURRENT FILING DATE: 2003-12-01
;  PRIOR APPLICATION NUMBER: US 60/461,133
;  PRIOR FILING DATE: 2003-04-09
;  NUMBER OF SEQ ID NOS: 1,067
;  SOFTWARE: Patentin version 3.0
;  SEQ ID NO 15
;  LENGTH: 675
;  TYPE: DNA
;  ORGANISM: Artificial
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(675)
;  OTHER INFORMATION: "mRFP1", an engineered monomeric form of DsRed FP
;  PUBLICATION INFORMATION:
;  AUTHORS: Campbell, R.E. et al
;  TITLE: A monomeric red fluorescent protein
;  JOURNAL: Proc. Nat'l. Acad. Sci.
;  VOLUME: 99
```

```
Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db  165 CACTGTTCCGTAAAGCTTATGTAACCAAAGGTGGACCTTTGCCATTTGCTTTTGATATT 224

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaGluPro 80
Db  225 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTAAAGCACCCTGCCGAACCA 284

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db  285 GACTATAAAAAGCTGTCATTTCCTGAGGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 344

Qy  101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db  345 GACGGTGGCGTGGTTACTGTAACCCAAGATTCCAGTTTGCAAGACGGCTGTTTCATCTAC 404

Qy  121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db  405 GAGGTCAAGTTCATTGGGGTGAACTTTCCTTCTGATGGACCCGTTATGCAGAAGAAGACA 464

Qy  141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspAspValLeuLysGlyGlu 160
Db  465 CGGGGCTGGGAAGCCAGCACGGAGCGTTTGTATCCTCGTGATGGCGTCCTGAAGGGCGAG 524

Qy  161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db  525 ATCCATATGGGCTCTGAGGCGGGAGGCGGCCATTACCTCGTTGAATTCAAAAGTATT 584

Qy  181 TyrMetAlaLysLysPro---ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeu 199
Db  585 TACATGGTAAAGAAGCCTTCAGTGCAGTTGCCAGGCTACTACTATTATGTTGACTCCAAACTG 644

Qy  200 AspIleThrSerHisAsnGluAspTyr......GluLysThrGlnGly 219
Db  645 GATATGACGAGCCACAACGAAGATTACACAGTCGTTGAGCAGTATGATGAAAAAACCCAGGGA 704

Qy  220 ArgHisHisLeuPheLeu 225
Db  705 CGCCACCATCCGTTCATT 722
```

; ISSUE: 12
; PAGES: 7877-82
; DATE: 2002-06-11
; RELEVANT RESIDUES: (1)..(675)
US-10-724-178-15

Alignment Scores:
Pred. No.:
Score:　　　　　　　　　　　　　　　　　　　　　1.56e-118
Percent Similarity:　　　　　　　　　　　　　　1021.00
Best Local Similarity:　　　　　　　　　　　　89.14%
Query Match:　　　　　　　　　　　　　　　　86.88%
DB:　　　　　　　　　　　　　　　　　　　　84.10%

Length:　675
Matches:　192
Conservative:　5
Mismatches:　24
Indels:　0
Gaps:　0

US-10-006-922A-12 (1-225) x US-10-724-178-15 (1-675)

```
Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValAlaArgMetGluGly  20
         |||||||||||||:::::|||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGGCGTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCGCATGGAGGGC  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GACACTAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  180

Qy   61  HisAsnThrValLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CACAACACCGTCAAGGTCACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  240

Qy   81  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC  300

Qy  101  LysValLysPheIleGlyValAlaAsnPheGlyProValMetGlnLysLysThr  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  AAGGTGAAGTTCATCGGCGTGGCTAACTTCGGCCCCGTAATGCAGAAGAAGACC  360

Qy  121  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ATGGGCTGGGAGGCCTCCACCGAGCGGGGATGTACCCGAGGACGGCGTGCTGAAGGGCGAG  420

Qy  141  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC  480

Qy  161  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC  540

Qy  181  IleThrSerHisAsnGluAspTyrThrIleValGluIleGluGlnTyrGluArg  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC  600

Qy  201  His  221
         |||
Db  661  CAC  663
```

; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER 2: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 678
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
;   OTHER INFORMATION: "mRPP1"
US-10-121-258-9

Alignment Scores:
Pred. No.:
Score:　　　　　　　　　　　　　　　　　　　1.57e-118
Percent Similarity:　　　　　　　　　　　　　1021.00
Best Local Similarity:　　　　　　　　　　　89.14%
Query Match:　　　　　　　　　　　　　　　86.88%
DB:　　　　　　　　　　　　　　　　　　　84.10%

Length:　678
Matches:　192
Conservative:　5
Mismatches:　24
Indels:　0
Gaps:　0

US-10-006-922A-12 (1-225) x US-10-121-258-9 (1-678)

```
Qy    1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValAlaArgMetGluGly  20
         |||||||||||||:::::|||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGGCGTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCGCATGGAGGGC  60

Qy   21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly  40
         :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC  120

Qy   41  AspThrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GACACTAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  180

Qy   61  HisAsnThrValLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CACAACACCGTCAAGGTCACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  240

Qy   81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
         :::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy  101  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC  360

Qy  121  LysValLysPheIleGlyValAlaAsnPheGlyProValMetGlnLysLysThr  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AAGGTGAAGTTCATCGGCGTGGCTAACTTCGGCCCCGTAATGCAGAAGAAGACC  420

Qy  141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
         |||||||||||||||||||||||::::||||||||||||||||||||||||||
Db  421  ATGGGCTGGGAGGCCTCCACCGAGCGGGGATGTACCCGAGGACGGCGTGCTGAAGGGCGAG  480

Qy  161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  ATCAAGATGAGGCTGAAGCTGAAAGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC  540

Qy  181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC  600

Qy  201  IleThrSerHisAsnGluAspTyrThrIleValGluIleGluGlnTyrGluArg  220
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC  660

Qy  221  His  221
         |||
Db  661  
```

RESULT 80
US-10-121-258-9
; Sequence 9, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Telen, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Telen, Roger
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME

RESULT 81
US-10-931-304-9
; Sequence 9, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
US-10-931-304-9
; OTHER INFORMATION: "mRFP1"

Alignment Scores:
Pred. No.:
Score: 1021.00
Percent Similarity: 89.14%
Best Local Similarity: 86.88%
Query Match: 84.10%

Length: 678
Matches: 192
Conservative: 5
Mismatches: 24
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-9 (1-678)

```
Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGGCCTCCTCGAGGAGGTCATCAAGGAGTTCATGGCTTCAAGGTGCGCATGGAGGGC 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy  61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy  81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACTTGAAGCTGTCCTTCCCGGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTGAAGCTCGCGCGCACCAACTTCCCTCGACGGCCCGTAATGCAGAAGAAGACC 420
```

RESULT 82
US-10-931-304-109
; Sequence 109, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 109
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mGrapel"
US-10-931-304-109

Alignment Scores:
Pred. No.:
Score: 1014.00
Percent Similarity: 89.14%
Best Local Similarity: 86.43%
Query Match: 83.53%

Length: 678
Matches: 191
Conservative: 6
Mismatches: 24
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-109 (1-678)

```
Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGGCCTCCTCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy  21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db  61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC 120

Qy  41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 ACCCAGACACGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180
```

Alignment Scores:

Pred. No.:
Score: 2.15e-117   Length: 678
Percent Similarity: 1012.00   Matches: 190
Best Local Similarity: 88.69%   Conservative: 6
                        85.97%   Mismatches: 25

RESULT 83
US-10-931-304-80
; Sequence 80, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "mRFP1.1"
US-10-931-304-80

```
Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIlePro   80
Db   181  CTGTCCCCTCAGTTCAGTACGGCTCCAAGGCCTACGTAGGTGAAGCACCCCGCCGACATCCC 240

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
Db   241  GACTACCTGAAGGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
Db   301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC   360

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
Db   361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC   420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
Db   421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG   480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
Db   481  ATCCACAAGGCGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC   540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
Db   541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTCGACTACAAGCTGGAC   600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
Db   601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC   660

Qy   221  His   221
Db   661  CAC   663
```

US-10-006-922A-12 (1-225) x US-10-931-304-80 (1-678)

```
Qy     1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly   20
Db     1  ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC   60

Qy    21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
Db    61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   120

Qy    41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db   121  ACCCAGACGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC  180

Qy    61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsnIlePro   80
Db   181  CTGTCCCCTCAGTTCAGTACGGCTCCAAGGCCTACGTAGGTGAAGCACCCCGCCGACATCCC 240

Qy    81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
Db   241  GACTACCTGAAGGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy   101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
Db   301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC   360

Qy   121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
Db   361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC   420

Qy   141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
Db   421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG   480

Qy   161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
Db   481  ATCCACAAGGCGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC   540

Qy   181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
Db   541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTCGACTACAAGCTGGAC   600

Qy   201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
Db   601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC   660

Qy   221  His   221
Db   661  CAC   663
```

RESULT 84
US-10-931-304-105
; Sequence 105, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308

; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 105
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mHoneydew"

US-10-931-304-105

US-10-006-922A-12 (1-225) x US-10-931-304-105 (1-678)

Alignment Scores:
Pred. No.:
Score: 9.2e-117
Percent Similarity: 1007.00
Best Local Similarity: 88.69%
Query Match: 85.07%
                         82.95%
                              9

                                    Length: 678
                                    Matches: 188
                                    Conservative: 8
                                    Mismatches: 25
                                    Indels: 0
                                    Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValAlaArgMetGluGly 20
Db   1 ATGGCCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGGCCAGGATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   181 CTGTCCCCTCAGTTCATGTGGGGCTCCAAGGGCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   241 GACTACTTGAAGCTGTCCTTCCCGGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   421 ATGGGCTGGGAGGCCAGCACCGAGCGGCTGTACCCCCGAGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   481 ATCCACAAGGCTCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC 540

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTATGTGGACAGCAAGCTGGAC 600

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC 660

Qy   221 His 221
Db   661 CAC 663

US-10-006-922A-12.p2nmlnscr.rnpbm

GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;   PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831C2P2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 101
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mTangerine"

US-10-931-304-101

US-10-006-922A-12 (1-225) x US-10-931-304-101 (1-678)

Alignment Scores:
Pred. No.:
Score: 3.94e-116
Percent Similarity: 1002.00
Best Local Similarity: 88.24%
Query Match: 85.52%
                         82.54%
                              9

                                    Length: 678
                                    Matches: 189
                                    Conservative: 6
                                    Mismatches: 26
                                    Indels: 0
                                    Gaps: 0

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValAlaArgMetGluGly 20
Db   1 ATGGCCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGGCCAGGATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   181 CTGTCCCCTCAGTTCTCTGTTACGGCTCCAAGGGCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   241 GACTACTTGAAGCTGTCCTTCCCGGAGGGCTTCAAGTGGGAGCGGGTGATGAACTTCGAG 300

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   361 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   421 ATGGGCTGGGAGGCCAGCACCGAGCGGCTGTACCCCCGAGACGGCGTGCTGAAGGGCGAG 480

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   481 ATCAAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC 540

RESULT 85
US-10-931-304-101
; Sequence 101, Application US/10931304
; Publication No. US20050196768A1

Qy 181 TyrMetAlaLeuLysProValGlnLeuProGlyTyrTyrTyrValAspSerIleLysLeuAsp 200
Db 541 TACATGGCCAAGAGACCGGTGCAGCTGCCCGGCTACTACTACGTCAGCATCAAGCTGGAC 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAATGTACGAGCGCACCGAAGGGCGC 660

Qy 221 His 221
Db 661 CAC 663

RESULT 86
US-10-931-304-86
; Sequence 86, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 86
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "OrS4-9"
US-10-931-304-86

Alignment Scores:
Pred. No.:
Score: 4.3e-115   Length: 711
                                         Matches: 186
Percent Similarity: 994.00   Conservative: 5
Best Local Similarity: 88.84%   Mismatches: 24
Query Match: 86.51%   Indels: 0
DB: 81.88%   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-86 (1-711)

Qy   7 ValIleLysGluGlyPheMetAlaArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu 26
Db   34 ATCATCAAGGAGAGTTCATGCGCTTCAAGGTACGCATGGAGGGCTCCGTGAACGGCCACGAG 93

Qy   27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
Db   94 TTCGAGATTGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAACACGGTGAAGCTG 153

Qy   47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
Db 154 AAGGTGACCAAGGGTGGGCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCCAC 213

Qy   67 TyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLysLeuSer 86
Db 214 TACGGCTCCAAGGCCTACGTGAAGCACCCCGCAGACATCCCCGACTACTTGAAGCTGTCC 273

Qy   87 PheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValValThr 106
Db 274 TTCCCCGAGGGCTTCAAGTGGGAACGGGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333

Qy 107 ValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIleGly 126
Db 334 GTGACACAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGC 393

Qy 127 ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer 146
Db 394 ACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC 453

Qy 147 ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLys 166
Db 454 TCCGAACGCCTGTACCCCCGAGGACGGCGTCCTGAAGGGCGAGATCCACAAGGCTCTGAAG 513

Qy 167 LeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
Db 514 CTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGAGCATCTACATGGCCAAGAAGCCC 573

Qy 187 ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu 206
Db 574 GTGCAGCTGCCCGGCTACTACTACGTCGACAGCAAGCTGGACATCACCTCCCACAACGAG 633

Qy 207 AspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221
Db 634 GACTACACCATCGTGGAACAGTACGAACGCACCGAAGGGCGCCAC 678

RESULT 87
US-10-931-304-90
; Sequence 90, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 90
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynuc. encoding polypeptide variant "mRFP" (P2Q6) (mGrape2)"
US-10-931-304-90

Alignment Scores:
Pred. No.:
Score: 7.69e-115   Length: 711
                                          Matches: 185
Percent Similarity: 992.00   Conservative: 6
Best Local Similarity: 88.84%   Mismatches: 24
Query Match: 86.05%   Indels: 0
DB: 81.71%   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-931-304-90 (1-711)

Qy   7 ValIleLysGluGlyPheMetAlaArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu 26
Db   34 GTCATCAAGGAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAG 93

Qy   27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
Db   94 TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAACACGGTGAAGCTG 153

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:

US-10-931-304-103

Length: 711
Matches: 185
Conservative: 5
Mismatches: 25
Indels: 0

1.84e-114
989.00
88.37%
86.05%
81.47%

GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  APPLICANT: Baird, Geoffrey
  TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
    PROTEIN VARIANTS AND METHODS FOR MAKING SAME
  FILE REFERENCE: 39754-0831CP2CP3
  CURRENT APPLICATION NUMBER: US/10/931,304
  CURRENT FILING DATE: 2004-08-30
  PRIOR APPLICATION NUMBER: 10/209,208
  PRIOR FILING DATE: 2002-07-29
  PRIOR APPLICATION NUMBER: 10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  NUMBER OF SEQ ID NOS: 110
  SEQ ID NO 103
  LENGTH: 711
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mOrange (mOPP1

RESULT 88
Sequence 103, Application US/10931304
Publication No. US20050196768A1

RESULT 89
Sequence 84, Application US/10931304
Publication No. US20050196768A1

GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  APPLICANT: Baird, Geoffrey
  TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
    PROTEIN VARIANTS AND METHODS FOR MAKING SAME
  FILE REFERENCE: 39754-0831CP2CP3
  CURRENT APPLICATION NUMBER: US/10/931,304
  CURRENT FILING DATE: 2004-08-30
  PRIOR APPLICATION NUMBER: 10/209,208
  PRIOR FILING DATE: 2002-07-29
  PRIOR APPLICATION NUMBER: 10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  NUMBER OF SEQ ID NOS: 110
  SEQ ID NO 84
  LENGTH: 711

RESULT 90

```
US-10-931-304-93
; Sequence 93, Application US200501967768A1
; Publication No. US20050196768A1
; GENERAL INFORMATION:
;  APPLICANT: Tsien, Roger
;  APPLICANT: Campbell, Robert
;  APPLICANT: Baird, Geoffrey
;  TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;  TITLE OF INVENTION:  PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;  FILE REFERENCE: 39754-0831CP2CP3
;  CURRENT APPLICATION NUMBER: US/10/931,304
;  CURRENT FILING DATE: 2004-08-30
;  PRIOR APPLICATION NUMBER: 10/209,208
;  PRIOR FILING DATE: 2002-07-29
;  PRIOR APPLICATION NUMBER: 10/121,258
;  PRIOR FILING DATE: 2002-04-10
;  PRIOR APPLICATION NUMBER: 09/866,538
;  PRIOR FILING DATE: 2001-05-24
;  PRIOR APPLICATION NUMBER: 09/794,308
;  PRIOR FILING DATE: 2001-02-26
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 93
;  LENGTH: 711
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mRFP2 (mCherry)
US-10-931-304-93

Alignment Scores:
Pred. No.:
Score:                  987.00
Percent Similarity:     88.84%
Best Local Similarity:  85.58%
Query Match:            9

                    Length:          711
                    Matches:         184
                    Conservative:    7
                    Mismatches:      24
                    Indels:          0
                    Gaps:            0
DB:
```

US-10-006-922A-12 (1-225) x US-10-931-304-93 (1-711)

```
Qy      7  ValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu  26
              ::||||||||||||||||||||||||:|||||||||||||||:::||||||
Db     34  ATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAG  93

Qy     27  PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu  46
               ||||||||||||||||||||||||||||||||||||||||||:|||||||
Db     94  TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAACACCGTGAAGCTG  153

Qy     47  LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln  66
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    154  AAGGTGACCAAGGGTGGGCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATG  213

Qy     67  TyrGlySerLysValTyrValLysHisProAlaAlaSerIleProAspTyrLysLysLeuSer  86
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db    214  TACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCC  273

Qy     87  PheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyGlyValThr  106
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    274  TTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC  333

Qy    107  ValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIleGly  126
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    334  GTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGC  393

Qy    127  ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer  146
              |||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db    394  ACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC  453

Qy    147  ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLys  166
              |||||||||||||||||||||||||||||||||||||||:||            ||:
Db    454  TCCGAGCGGCTGTACCCCCGAGGACGGCGTGCTGAAGGGCGAGATCAAGCAGAGGCTGAAG  513

Qy    167  LeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro  186
              ||||||:||||||||||||||||||||||||||||||  ::
Db    514  CTGAAGGACGGCGGCCACTACCTGGTCGAGTTCAAGAGCATCTACATGGCCAAGAAGCCC  573

Qy    187  ValGlnLeuProGlyTyrTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu  206
              |||||||||||||||||||||||||||||||:||||||||||||||||||
Db    574  GTGCAGCTGCCCGGCTACTACTACTACGTGGACAGCAAGCTGGACATCACCTCCCACAACGAG  633

Qy    207  AspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis  221
              |||||||||||||||||||||||||||||||||||||||||
Db    634  GACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCAC  678
```

RESULT 91

US-10-931-304-95
; Sequence 95, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SEQ ID NO 95
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "myOFP (74-11)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 185
; OTHER INFORMATION: n = A,T,C or G
US-10-931-304-95

; SOFTWARE: FastSEQ for Windows Version 4.0

Alignment Scores:
Pred. No.:
Score: 4.37e-114
Percent Similarity: 986.00
Best Local Similarity: 87.91%
Query Match: 85.58%
DB: 81.22%
                          9

US-10-006-922A-12 (1-225) x US-10-931-304-95 (1-708)

Length: 708
Matches: 184
Conservative: 5
Mismatches: 26
Indels: 0
Gaps: 0

Qy   7 ValIleLysGlyGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu 26
            ::|||||||||||||||||||||||||||||||:||||||||||||||||||||||::|||
Db   34 AtCatCAAGGAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCACGGAG 93

Qy   27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
            ||:||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db   94 TTCGAGATCGAGGGCGAGGGCGAGGGCCGGCCCTACGAGGGCTTTCAGACCGGCTAAGTCTG 153

Qy   47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   154 AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCNTCGGACATCCTGTCCCCTCAGTTCACC 213

Qy   154 ...
Db   274 TTCCCCGAGGGCTTCAAGGTGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333

Qy   87 PheProGluGlyPheLysValGluArgValMetAsnPheGluAspGlyGlyValValThr 106
            ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db   274 TTCCCCGAGGGCTTCAAGGTGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333

Qy   107 ValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIleGly 126
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   334 GTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGC 393

Qy   127 ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer 146
            |||||||||||||||||||||||||||||||:|||||||||||||||||||
Db   394 ACCAACTTCCCTCCCGACGGGCCCGTNATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC 453

Qy   147 ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLys 166
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   454 TCCGAGCGCCTGTACCCCCGAGGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAG 513

RESULT 92
US-10-931-304-97
; Sequence 97, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SEQ ID NO 97
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mROFP (A2/6-6)"
US-10-931-304-97

; SOFTWARE: FastSEQ for Windows Version 4.0

Alignment Scores:
Pred. No.:
Score: 1.93e-112
Percent Similarity: 973.00
Best Local Similarity: 87.44%
Query Match: 84.65%
DB: 80.15%
                          9

US-10-006-922A-12 (1-225) x US-10-931-304-97 (1-711)

Length: 711
Matches: 182
Conservative: 6
Mismatches: 27
Indels: 0
Gaps: 0

Qy   7 ValIleLysGlyGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu 26
            ::|||||||||||||||||||||||||||||||:||||||||||||||||||||::|||
Db   34 AtCatCAAGGAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGGCTCCGTGAACGGCCACGAG 93

Qy   27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
            ||:||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db   94 TTCGAGATCGAGGGCGAGGGCGAGGGCCGGCCCTACGAGGGCCACGCCAAGCTG 153

Qy   47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
            |||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db   154 AAGGTGACCAAGGGTGGCCCCCTTCGCCCTTCGCAaAspIle 213

Qy   27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
            ||:||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db   94 TTCGAGATCGAGGGCGAGGGCGAGGGCCGGCCCTACGAGGGCCACGCCAAGCTG 153

Qy   67 TyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLysLeuSer 86
            ||||||||||||||||||||||:|||aAsp||||||||||||:||||||||
Db   214 TACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCC 273

Qy   47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
            |||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db   154 AAGGTGACCAAGGGTGGCCCCCTTCGCCCTTCGCA 213

Qy   87 PheProGluGlyPheLysValGluArgValMetAsnPheGluAspGlyGlyValValThr 106
            ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||

RESULT 93
US-10-931-304-99
; Sequence 99, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant "mStrawberry"
US-10-931-304-99

Alignment Scores:
Pred. No.:
Score: 973.00   Length: 711
Percent Similarity: 87.44%   Matches: 182
Best Local Similarity: 84.65%   Conservative: 6
Query Match: 80.15%   Mismatches: 27
DB: Indels: 0
                                         Gaps: 0

US-10-006-922a-12 (1-225) x US-10-931-304-99 (1-711)

```
Qy  274 TTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333
        |||||
Db  107 ValThrGlnAspSerLeuGlnAspGlyCysPheIleTyrLysValLysSerPheIleGly 126

Qy  334 GTGACCCAGGACTCTCCCCTGCCGAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGC 393
        |||||
Db  127 ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer 146

Qy  394 ACCAACTTCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC 453
        |||||
Db  147 ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLys 166

Qy  454 TCCGAGGACGGATGTACCCCGAGGACGGCGAGATCAAGATGAGGGCTGAAG 513
        ::|||
Db  167 LeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186

Qy  514 CTGAAGGACGGGGGCCACTACCTGGTCGAGTTCAAGAGCATCTACATGGCCAAGAAGCCC 573
        |:::
Db  187 ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu 206

Qy  574 GTGCAGCTGCCCGGCTACTACTACGTCGACAGCAAGTTGGACATCACCTCCCACAACGAG 633
        |||||
Db  207 AspTyrTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221

Qy  634 GACTACTACACCATCGTGGAACTGTACGAACGGCGCCGGAGGCGCCGCCAC 678
```

RESULT 94
US-10-931-304-88
; Sequence 88, Application US/10931304
; Publication No. US20050196768A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: 39754-0831CP2CP3
; CURRENT APPLICATION NUMBER: US/10/931,304
; CURRENT FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: 10/209,208
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 88
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide of polypeptide variant "Y1.3 (mYOFP 1.3) (mBanana)"
US-10-931-304-88

Alignment Scores:
Pred. No.:
Score: 960.00   Length: 711
                         Matches: 180

```
Qy   94 TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAAGGCACCCAGACCGCCAAGCTG 153
        |||||
Db   47 LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66

Qy  154 AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTAACCCCCCAGTTCACC 213
        |||||
Db   67 TyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLysLeuSer 86

Qy  214 TACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCC 273
        |||||
Db   87 PheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValValThr 106

Qy  274 TTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333
        |||||
Db  107 ValThrGlnAspSerLeuGlnAspGlyCysPheIleTyrLysValLysSerPheIleGly 126

Qy  334 GTGACCCAGGACTCTCCCCTGCCGAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGC 393
        |||||
Db  127 ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer 146

Qy  394 ACCAACTTCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC 453
        |||||
Db  147 ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLys 166

Qy  454 TCCGAGGACGGATGTACCCCGAGGACGGCGAGATCAAGATGAGGGCTGAAG 513
        |::||
Db  167 LeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186

Qy  514 CTGAAGGACGGGGGCCACTACCTGGTCGAGTTCAAGAGCATCTACATGGCCAAGAAGCCC 573
        :::|
Db  187 ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu 206

Qy  574 GTGCAGCTGCCCGGCTACTACTACGTCGACAGCAAGTTGGACATCACCTCCCACAACGAG 633
        |||||
Db  207 AspTyrTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221

Qy  634 GACTACTACACCATCGTGGAACTGTACGAACGGCGCCGGAGGCGCCGCCAC 678
```

US-10-931-304-99

Alignment Scores:
Pred. No.:
Score: 8.42e-111   Length: 711
                            Matches: 180

```
Qy    7 ValIleLeuLysGluGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu 26
        ||:|||||||||||||||:||||||||||||:|||
Db   34 ATCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTGAACGGCCACGAG 93

Qy   27 PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
```

```
Percent Similarity:    86.05%        Conservative:    5
Best Local Similarity: 83.72%        Mismatches:     30
Query Match:           79.08%        Indels:          0
DB:                        9         Gaps:            0

US-10-006-922A-12 (1-225) x US-10-931-304-88 (1-711)

Qy    7  ValIleLysGluPheMetArgPheLysValArgMetGluGlyThrValAsnGlyHisGlu 26
            ::   |    |  |      | |  | :|:||||    |    |      |    |
Db   34  GTCATCAAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCCTCCGTGAACGGCACGAG 93

Qy   27  PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsnThrValLysLeu 46
            ||  |  ||      | ||   ||      |   |   ||   |   |      |
Db   94  TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCCAAGCTG 153

Qy   47  LysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPheGln 66
            ||      |      |     |  ||   |      | ||     |  |    |  |
Db  154  AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCTGC 213

Qy   67  TyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyrLysLysLeuSer 86
            | ||  |      |     |   |  |   |   |    |   |   |      |
Db  214  TACGGCTCCAAGGGCTACGTGAAGCACCCCGCTGATCCCCGACTACTTCAAGCTGTCC 273

Qy   87  PheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValValThr 106
            ||      |  |   |  |      |   |   |   |  |      |      | |
Db  274  TTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC 333

Qy  107  ValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIleGly 126
            | |  |      |      |      || || | |      |    |    ||
Db  334  GTGGCTCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGC 393

Qy  127  ValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSer 146
            |     |  |      |   |    |   |  |      |      |      |
Db  394  ACCAACTTCCCCTCCGACGGGCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC 453

Qy  147  ThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLys 166
            | ||  |  |   |   |      | | |   ||      | |      |
Db  454  TCCGACCGGGGATGTACCCCCGAGGACGGCGTGCTGAAGGGCGAGATCAAGAACTTCAAG 513

Qy  167  LeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
            |  |   |   | |   |   | ||   |      |      |   |
Db  514  CTGAAGGACGGCGGCCACTACAGCGCGGCGAGACCAAGGCCAACCTACAAGGCCC 573

Qy  187  ValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThrSerHisAsnGlu 206
            | || |   |      |         |   |      | |   |  |      |
Db  574  GTGCAGTTGCCCGGCCGCTACTATAGCGGCGGAAGATCGACATCACCTCCCACAATGAG 633

Qy  207  AspTyrThrIleValGluGlnTyrGluArgThrGluGlyArgHis 221
            | |      |   |      |    |      |  | | |
Db  634  GACTACACTATCGTGGAAATTGTACGGCGCGGCGGCCAC 678
```

```
FEATURE:
OTHER INFORMATION: modif. frag.; RFP F1F, with position 1 Met removed
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(549)
OTHER INFORMATION: RFP F1F  corresponds to aa residues 1-184 of mRFP

US-10-724-178-1040

Alignment Scores:
Pred. No.:                           Length:        549
Score:             848.00            Matches:       159
Percent Similarity: 90.11%           Conservative:    5
Best Local Similarity: 87.36%        Mismatches:     18
Query Match:       69.85%            Indels:          0
DB:                        7         Gaps:            0

US-10-006-922A-12 (1-225) x US-10-724-178-1040 (1-549)

Qy    3  SerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
         |   |:::|:::|||||||||||||||||||||||||||||||||||||||||:::|||
Db    4  TCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCCTCGTG 63

Qy   23  AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64  AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAAC 123

Qy   43  ThrValLysValLysLeuGlyValThrLysGlyGlyProLeuProPheAlaTrpAspIle 62
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124  ACCGTCAAGGTCAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 183

Qy   63  ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db  184  CCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTAC 243

Qy   83  LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  244  TTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 303

Qy  103  GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal 122
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  304  GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG 363

Qy  123  LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  364  AAGCTGCCGGCCAGCAAACTTCCCTCCGACGGGCCCGTAATGCAGAAGAAGACCATGGGC 423

Qy  143  TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis 162
         |:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  424  TGGGAGGCCTCCACCGAGCGGACCGAGGGCGCCCTGAAGGGCGAGATCAAG 483

Qy  163  LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet 182
         ||||||||||||||||||||||||||||||||||||||:::||
Db  484  ATGAGGGTGAAGCTGAAGGACGGCGGCCACTACGAGGTCAAGAACCACACCTACATG 543

Qy  183  AlaLys 184
         ||||||
Db  544  GCCAAG 549
```

```
Search completed: January 12, 2006, 15:16:29
Job time : 837 secs
```

US-10-006-922A-12

```
Title:
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..............EDYIVEQYRTEGRHHLFL 225

Scoring table:   BLOSUM62
                 Xgapop   10.0 ;  Xgapext   0.5
                 Ygapop   10.0 ;  Ygapext   0.5
                 Fgapop    6.0 ;  Fgapext   7.0
                 Delop     6.0 ;  Delext    7.0

Searched:   6038814 seqs, 404674181 residues

Total number of hits satisfying chosen parameters:        15
```

OM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2006 Compugen Ltd.

GenCore version 5.1.6

Run on: January 12, 2006, 12:55:43 ; Search time 381 Seconds
                                       477.962 Million cell updates/sec
                                       (without alignments)

Searched: US-10-006-922A-12

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 69%
                 Maximum Match 100%
                 Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_P/US10006922/runat_10012006_162506_19200/app_query.fasta_1.391
-DB=Published_Applications_NA_New -QFMT=fastap -SUFFIX=p2nminscr.rnpbn
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bts -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=500 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=69 -ALIGN=500 -MODE=LOCAL -OUTFMT=pto -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_1_184 @runat_10012006_162506_19200 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA_New:*
             1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
             2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
             3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
             4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
             5: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
             6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
             7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
             8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
             9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
            10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
```
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

```
Result            %     Query
No.   Score  Match  Length  DB  ID                  Description
----  -----  -----  ------  --  -----------------   -----------
 1    1214   100.0    678    6  US-10-209-208-2     Sequence 2, Appli
 2    1214   100.0    678    7  US-11-218-880-2     Sequence 5, Appli
 3    1210    99.7    681    6  US-10-209-208-3     Sequence 3, Appli
 4    1210    99.7    681    6  US-10-209-208-23    Sequence 23, Appl
 5    1210    99.7    681    7  US-11-218-880-23    Sequence 23, Appl
 6    1210    99.7    681    7  US-11-218-880-23    Sequence 3, Appl
 7    1186    97.7    6706   6  US-10-655-872-4     Sequence 4, Appl
```

US-10-006-922A-12

```
Title:
Perfect score: 1214
Sequence: 1 MRSSKNVIKEFMRFKVRMEG..............EDYIVEQYRTEGRHHLFL 225

Scoring table:   BLOSUM62
                 Xgapop   10.0 ;  Xgapext   0.5
                 Ygapop   10.0 ;  Ygapext   0.5
                 Fgapop    6.0 ;  Fgapext   7.0
                 Delop     6.0 ;  Delext    7.0

Searched:   6038814 seqs, 404674181 residues

Total number of hits satisfying chosen parameters:        15
```

OM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2006 Compugen Ltd.

GenCore version 5.1.6

Run on: January 12, 2006, 12:55:43 ; Search time 381 Seconds
                                       477.962 Million cell updates/sec
                                       (without alignments)

Searched: US-10-006-922A-12

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 69%
                 Maximum Match 100%
                 Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_P/US10006922/runat_10012006_162506_19200/app_query.fasta_1.391
-DB=Published_Applications_NA_New -QFMT=fastap -SUFFIX=p2nminscr.rnpbn
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bts -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=500 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=69 -ALIGN=500 -MODE=LOCAL -OUTFMT=pto -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10006922 @CGN_1_1_184 @runat_10012006_162506_19200 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA_New:*
             1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
             2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
             3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
             4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
             5: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
             6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
             7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
             8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
             9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
            10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result            %     Query
No.   Score  Match  Length  DB  ID                  Description
----  -----  -----  ------  --  -----------------   -----------
 1    1214   100.0    678    6  US-10-209-208-2     Sequence 2, Appli
 2    1214   100.0    678    7  US-11-218-880-2     Sequence 5, Appli
 3    1210    99.7    681    6  US-10-209-208-3     Sequence 3, Appli
 4    1210    99.7    681    6  US-10-209-208-23    Sequence 23, Appl
 5    1210    99.7    681    7  US-11-218-880-23    Sequence 23, Appl
 6    1210    99.7    681    7  US-11-218-880-23    Sequence 3, Appl
 7    1186    97.7    6706   6  US-10-655-872-4     Sequence 4, Appl

  8    1186   97.7   7927   6  US-10-655-872-7     Sequence 7, Appli
  9    1160   95.6    678   6  US-10-209-208-5     Sequence 5, Appli
 10    1160   95.6    678   7  US-11-218-880-5     Sequence 5, Appli
 11    1121   92.3    681   6  US-10-209-208-7     Sequence 7, Appli
 12    1121   92.3    681   7  US-11-218-880-7     Sequence 7, Appli
 13    1021   84.1    678   6  US-10-209-208-9     Sequence 9, Appli
 14    1021   84.1    678   7  US-11-218-880-9     Sequence 9, Appli
 15    1012   83.4    678   6  US-10-209-208-80    Sequence 80, Appl
```

ALIGNMENTS

RESULT 1
US-10-209-208-2

; Sequence 2, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 678
;   TYPE: DNA
;   ORGANISM: Discosoma sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(678)
;   OTHER INFORMATION: wild-type DsRed
US-10-209-208-2

Alignment Scores:
Pred. No.:
Score: 1214.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 6

                          Length: 678
                          Matches: 225
                          Conservative: 0
                          Mismatches: 0
                          Indels: 0
                          Gaps: 0

Length: 2.23e-133

US-10-006-922A-12 (1-225) x US-10-209-208-2 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGTCAATGGCCACGAGTTTGAAATAGAAGGCGAAGGAGAGGGCCATACCGAAGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTGTCACCACAAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCGACATACCA 240

Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300

Qy 101 AspGlyLysGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGGTGGGGTCGTTACTGTAACCCAGGATTCCAGTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisHisLeuPheLeu 225
         ||||||||||||||||
Db 661 CACCATCTGTTCCTT 675

RESULT 2
US-11-218-880-2

; Sequence 2, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(678)
; OTHER INFORMATION: wild-type DsRed
US-11-218-880-2

Alignment Scores:
Pred. No.:
Score: 2.23e-133   Length: 678
                          1214.00   Matches: 225
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
                          7   Gaps: 0
DB:

US-10-006-922A-12 (1-225) x US-11-218-880-2 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40

Db 61 AGGGTCAATGGGCACCAGTTCGAGATTGAAGGCGAAGGTGAAGGCCCATACGAAGGC 120

Db 41 HisHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
        CACAATACCGTAAAGCTGAAGGTAACCAAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaIlePro 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TTGTCACCACAATTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240

Qy 81 AspGlyTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTTCATGAACTTTGAA 300

Qy 101 AspGlyLysGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GACGGGTGGGGTCGTTACTGTAACCCAGGATTCCAGTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AAGGTCAAGTTCATTGGCGTGAACTTTCCTTCCGATGGACCTGTTATGCAAAAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACATGGCAAAGAAGCCTGTCCAGCTACCAGGGTACTACTATGTTGACTCCAAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660

Qy 221 HisHisHisLeuPheLeu 225
        ||||||||||||||||
Db 661 CACCATCTGTTCCTT 675

RESULT 3
US-10-209-208-3

; Sequence 3, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-10-209-208-3

; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-10-209-208-23

Alignment Scores:
Pred. No.:
Score: 6.67e-133
Percent Similarity: 1210.00   Length: 681
Best Local Similarity: 100.00%   Matches: 224
Query Match: 99.56%   Conservative: 1
                               6   Mismatches: 0
                                                 Indels: 0
                                                 Gaps: 0
DB:

US-10-006-922A-12 (1-225) x US-10-209-208-23 (1-681)

(sequence alignment data)

RESULT 4
US-10-209-208-23
; Sequence 23, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 10/121,258
; PRIOR FILING DATE: 2002-04-10

Alignment Scores:
Pred. No.:
Score: 6.67e-133
Percent Similarity: 1210.00   Length: 681
Best Local Similarity: 100.00%   Matches: 224
Query Match: 99.56%   Conservative: 1
                               6   Mismatches: 0
                                                 Indels: 0
                                                 Gaps: 0
DB:

US-10-006-922A-12 (1-225) x US-10-209-208-3 (1-681)

RESULT 5
US-11-218-880-3
; Sequence 3, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-11-218-880-3

Alignment Scores:
Pred. No.:
Score: 6.67e-133   Length: 681
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
DB: 99.67%   Indels: 0
                                       Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-3 (1-681)

Qy   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
          :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCTTCCTCCAAGAACGTCATCAAGGAGTTCATGCGTTCAAGGTGCGCATGGAGGGC   63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   243

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   303

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC   363

Qy   121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr   140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC   423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu   160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG   483

Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC   543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp   200
          ||:||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC   603

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg   220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC   663

Qy   221 HisHisLeuPheLeu   225
          |||||||||||||||
Db   664 CACCACCTGTTCCTG   678

RESULT 6
US-11-218-880-23
; Sequence 23, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-11-218-880-23

Alignment Scores:
Pred. No.:
Score: 6.67e-133   Length: 681
Percent Similarity: 1210.00   Matches: 224
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.56%   Mismatches: 0
DB: 99.67%   Indels: 0
                                       Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-23 (1-681)

Qy   1 MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly   20
          :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GTGCGCTTCCTCCAAGAACGTCATCAAGGAGTTCATGCGTTCAAGGTGCGCATGGAGGGC   63

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC   123

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC   183

Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   243

Qy   81 AspThrTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu   100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG   303

Left column

```
Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 GACGGGGGGGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 363

Qy   121 LysValLysPheIleGlyLysValAsnPheProSerAspGlyGlyProValMetGlnLysGlyThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCGGCCCCGTAATGCAGAAGAGAACC 423

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyGlyProValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGGCCCCGTGCTGAAGGGCGAG 483

Qy   161 IleHisLysAlaLeuLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 ATCCACAAGGCCCTGCTGAAGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 543

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 603

Qy   201 IleThrSerHisAlaAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 663

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db   664 CACCACCTGTTCCTG 678
```

RESULT 7
US-10-655-872-4

```
;   Sequence 4, Application US/10655872
;   Publication No. US20050251872A1
;   GENERAL INFORMATION:
;   APPLICANT: Bear, et al.
;   TITLE OF INVENTION: Lentiviral Vectors, Related Reagents, and Methods of Use Thereof
;   FILE REFERENCE: 0492611-0512
;   CURRENT APPLICATION NUMBER: US/10/655,872
;   CURRENT FILING DATE: 2003-09-05
;   SOFTWARE: PatentIn version 3.2
;   NUMBER OF SEQ ID NOS: 39
;   SEQ ID NO 4
;   LENGTH: 6706
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: lentiviral vector sequence
US-10-655-872-4
```

Alignment Scores:

Pred. No.:	1e-128	Length: 6706
Score:	1186.00	Matches: 219
Percent Similarity:	98.22%	Conservative: 2
Best Local Similarity:	97.33%	Mismatches: 4
Query Match:	97.69%	Indels: 0
DB:		Gaps: 0

US-10-006-922A-12 (1-225) x US-10-655-872-4 (1-6706)

```
Qy     1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2778 ATGGCCCTCCTCCGAGAACCGTTCATCACCGAGTTCATGGCCTTCAAGGTGCCCATGGAGGGC 2837

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyTyrProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2838 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCTACCCCTACGAGGGC 2897

Qy    41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2898 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 2957

Qy    61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2958 CTGTCCCCGCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 3017
```

Right column

```
Qy    81 AspThrTrpLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||
Db  3018 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 3077

Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3078 GACGGCGGCGTGGTCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 3137

Qy   121 LysValLysPheIleGlyLysValAsnPheProSerAspGlyGlyProValMetGlnLysGlyThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3138 AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCGGCCCCGTGATGCAGAAGAGAACC 3197

Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyGlyProValLeuLysGlyGlu 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3198 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGGCCCCGTGCTGAAGGGCGAG 3257

Qy   161 IleHisLysAlaLeuLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3258 ATCCACAAGGCCCTGCTGAAGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 3317

Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
         |||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db  3318 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 3377

Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3378 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 3437

Qy   221 HisHisLeuPheLeu 225
         |||||||||||||||
Db  3438 CACCACCTGTTCCTG 3452
```

RESULT 8
US-10-655-872-7

```
;   Sequence 7, Application US/10655872
;   Publication No. US20050251872A1
;   GENERAL INFORMATION:
;   APPLICANT: Bear, et al.
;   TITLE OF INVENTION: Lentiviral Vectors, Related Reagents, and Methods of Use Thereof
;   FILE REFERENCE: 0492611-0512
;   CURRENT APPLICATION NUMBER: US/10/655,872
;   CURRENT FILING DATE: 2003-09-05
;   SOFTWARE: PatentIn version 3.2
;   NUMBER OF SEQ ID NOS: 39
;   SEQ ID NO 7
;   LENGTH: 7927
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: lentiviral vector sequence
US-10-655-872-7
```

Alignment Scores:

Pred. No.:	1.26e-128	Length: 7927
Score:	1186.00	Matches: 219
Percent Similarity:	98.22%	Conservative: 2
Best Local Similarity:	97.33%	Mismatches: 4
Query Match:	97.69%	Indels: 0
DB:		Gaps: 0

US-10-006-922A-12 (1-225) x US-10-655-872-7 (1-7927)

```
Qy     1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
         ||||||||||||||||||||||||||||||||||||||:::||||
Db  3999 ATGGCCCTCCTCCGAGAACCGTTCATCACCGAGTTCATGGCCTTCAAGGTGCCCATGGAGGGC 4058

Qy    21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyTyrProTyrGluGly 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4059 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCTACCCCTACGAGGGC 4118

Qy    41 HisAsnThrValLysLeuLysValThrLysGly 60
         ||||||||||||||||||||||||||||||||||||
```

Alignment Scores:
```
Pred. No.:
Score:                      5.3e-127         Length:           678
Percent Similarity:         1160.00          Matches:          216
Rest Local Similarity:      97.33%           Conservative:     3
                            96.00%           Mismatches:       6
```

US-10-209-208-5

RESULT 9
US-10-209-208-5
Sequence 5, Application US/10209208
Publication No. US20050244921A1
GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  APPLICANT: Geoffrey Baird
  TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
  TITLE OF INVENTION: FOR MAKING SAME
  FILE REFERENCE: UC083.1CP2CP2
  CURRENT APPLICATION NUMBER: US/10/209,208
  CURRENT FILING DATE: 2002-07-29
  PRIOR APPLICATION NUMBER: 10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  NUMBER OF SEQ ID NOS: 80
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 5
  LENGTH: 678
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION:
  OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
  US-10-209-208-5

Qy   4119 CACAACACCGTGAAGCTGACGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 4178
Db   4179 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCGACATCCC 4238
Db   4239 GACTACAAGAAGGTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 4298
Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlySerPheIleTyr 120
Db   4299 GACGGCGGCGTGGGCACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 4358
Qy   121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysThr 140
Db   4359 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGACC 4418
Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   4419 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 4478
Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   4479 ACCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 4538
Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   4539 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 4598
Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnInTyrGluArgThrGluGlyArg 220
Db   4599 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 4658
Qy   221 HisHisHisLeuPheLeu 225
Db   4659 CACCACCTGTTCCTG 4673

Query Match: 95.55%   Indels: 0
DB: 6   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-5 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db   1 ATGGGCCGCTCCTCCGAGAACGTCATCAAGGAGTTCATGCGGTTCAAGGTGCGCATGGAG 60
Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120
Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db   121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180
Qy   61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db   181 CTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240
Qy   81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db   241 GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300
Qy   101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360
Qy   121 LysValLysPheIleGlyValAlaAsnPheProSerAspGlyProValMetGlnLysThr 140
Db   361 AAGGTGAAGTTCATCGGCGTGGCCAACTTCCCCTCCGACGGCCCCGTGATGCAGAAGACT 420
Qy   141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG 480
Qy   161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC 540
Qy   181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600
Qy   201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db   601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660
Qy   221 HisHisHisLeuPheLeu 225
Db   661 CACCACCTGTTCCTG 675

RESULT 10
US-11-218-880-5
Sequence 5, Application US/11218880
Publication No. US20060003420A1
GENERAL INFORMATION:
  APPLICANT: Tsien, Roger
  APPLICANT: Campbell, Robert
  TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
  TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
  FILE REFERENCE: UC083.1CP2CP1
  CURRENT APPLICATION NUMBER: US/11/218,880
  CURRENT FILING DATE: 2005-09-01
  PRIOR APPLICATION NUMBER: US/10/121,258
  PRIOR FILING DATE: 2002-04-10
  PRIOR APPLICATION NUMBER: 09/794,308
  PRIOR FILING DATE: 2001-02-26
  PRIOR APPLICATION NUMBER: 09/866,538
  PRIOR FILING DATE: 2001-05-24
  NUMBER OF SEQ ID NOS: 78
  SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "T1"
US-11-218-880-5

Alignment Scores:
Pred. No.:
Score: 5.3e-127
Percent Similarity: 1160.00
Best Local Similarity: 97.33%
Query Match: 96.00%
DB: 95.55%

Length: 678
Matches: 216
Conservative: 3
Mismatches: 6
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-5 (1-678)

```
Qy   1  MetArgSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyGly   20
Db   1  ATGGCCTCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC   60

Qy  21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly   40
Db  61  TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC   120

Qy  41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle   60
Db 121  ACCCAGACGCCCAAGCTGAAGGTGACCAAGGGCGGGCCCCCGCCCTTCGCCTGGGACATC   180

Qy  61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro   80
Db 181  CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC   240

Qy  81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu  100
Db 241  GACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG  300

Qy 101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr  120
Db 301  GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC  360

Qy 121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr  140
Db 361  AAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT  420

Qy 141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu  160
Db 421  ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAG  480

Qy 161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle  180
Db 481  ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC  540

Qy 181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp  200
Db 541  TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC  600

Qy 201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg  220
Db 601  ATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC  660

Qy 221  HisHisLeuPheLeu  225
Db 661  CACCACCTGTTCCTG  675
```

APPLICANT: Campbell, Robert
APPLICANT: Geoffrey Baird

TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
TITLE OF INVENTION: FOR MAKING SAME
FILE REFERENCE: UC083.1CP2CP2
CURRENT APPLICATION NUMBER: US/10/209,208
CURRENT FILING DATE:
PRIOR APPLICATION NUMBER: 2002-07-29
PRIOR FILING DATE: 10/121,258
PRIOR APPLICATION NUMBER: 2002-04-10
PRIOR FILING DATE: 09/866,538
PRIOR APPLICATION NUMBER: 2001-05-24
PRIOR FILING DATE: 09/794,308
PRIOR APPLICATION NUMBER: 2001-02-26
NUMBER OF SEQ ID NOS: 80
SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 7
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "dimer2"
US-10-209-208-7

Alignment Scores:
Pred. No.:
Score: 2.15e-122
Percent Similarity: 1121.00
Best Local Similarity: 95.96%
Query Match: 92.83%
DB: 92.34%

Length: 681
Matches: 207
Conservative: 7
Mismatches: 9
Indels: 0
Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-7 (1-681)

```
Qy   3  SerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyThrVal   22
Db  10  TCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTG   69

Qy  23  AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn   42
Db  70  AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAAC  129

Qy  43  ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer   62
Db 130  ACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC  189

Qy  63  ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr   82
Db 190  CCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTAC  249

Qy  83  LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly  102
Db 250  AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC  309

Qy 103  GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal  122
Db 310  GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG  369

Qy 123  LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly  142
Db 370  AAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC  429

Qy 143  TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis  162
Db 430  TGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC  489

Qy 163  LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet  182
Db 490  AAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATG  549

Qy 183  AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr  202
Db 550  GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACC  609
```

; GENERAL INFORMATION:
APPLICANT: Tsien, Roger

RESULT 11
US-10-209-208-7
; Sequence 7, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
APPLICANT: Tsien, Roger

```
Qy  203 SerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArgHisHis 222
            ||||||||||||||||||||||||||::::|||||||::||||||
Db  610 TCCCACAACGAGGACTACACCATCGTGGAACAGACAGTACGAGGGCCGCCACCAC 669

Qy  223 LeuPheLeu 225
            |||||||||
Db  670 CTGTTCCTG 678
```

RESULT 12
US-11-218-880-7
; Sequence 7, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "dimer2"
US-11-218-880-7

Alignment Scores:
Pred. No.:
Score: 2.15e-122   Length: 681
                    1121.00   Matches: 207
Percent Similarity: 95.96%   Conservative: 7
Best Local Similarity: 92.83%   Mismatches: 9
Query Match: 92.34%   Indels: 0
DB: 7   Gaps: 0

US-10-006-922A-12 (1-225) x US-11-218-880-7 (1-681)

Qy   3 SerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
            ||:::||||||||||||||||||||||||||||||||::||||
Db   10 TCCTCCGGAACGTCATCAAAGAGTTCATGCGCTTCAAAGGTGCCGGATGGGCTCCGTG 69

Qy   23 AsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyHisAsn 42
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   70 AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACCAG 129

Qy   43 ThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 130 ACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC 189

Qy   63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
            ||||||||||||||||||||||||||||||||||||||||||||||
Db 190 CCCCAGTTCCAGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGACTAC 249

Qy   83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
            |||||||||||||||||||||||||||||||||||||||||||||||
Db 250 AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGC 309

Qy 103 GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal 122
            |||||||||||||||||||||||||||||||||||||||||||
Db 310 GGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGGTGCTTCATCTACAAGGTG 369

Qy 123 LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
```

```
Db  370 AAGTTCGGCGGCGCACCAACTTCCCCAGCGACGGCCCCGTAATGCAGAAGAAGACCATGGGC 429
            ||||||||||||||||||||||||||||||||||||||||||
Qy  143 TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGluIleHis 162
            |||||||||||||||||||||||||||||||||||||||||||
Db  430 TGGGAGGCCTCCACCGAGCGGCTGTACCCCCGGGACGGCGTGCTGAAGGGCGAGATCCAC 489

Qy  163 LysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIleTyrMet 182
            ||||||||||||||||||||||||||||||||||||||||||||
Db  490 CAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATCTACATG 549

Qy  183 AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr 202
            ||||||||||||||||||||||||||||||||||||||||||
Db  550 GCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGACATCACC 609

Qy  203 SerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArgHisHis 222
            ||||||||||||||||||||||||||:::||||||::|||||
Db  610 TCCCACAACGAGGACTACACCATCGTGCAGTACGAGCGCACCGAGGGCCGCCACCAC 669

Qy  223 LeuPheLeu 225
            |||||||||
Db  670 CTGTTCCTG 678
```

RESULT 13
US-10-209-208-9
; Sequence 9, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
; OTHER INFORMATION: "mRFP1"
US-10-209-208-9

Alignment Scores:
Pred. No.:
Score: 1.36e-110   Length: 678
                    1021.00   Matches: 192
Percent Similarity: 89.14%   Conservative: 5
Best Local Similarity: 86.88%   Mismatches: 24
Query Match: 84.10%   Indels: 0
DB: 6   Gaps: 0

US-10-006-922A-12 (1-225) x US-10-209-208-9 (1-678)

Qy   1 MetArgSerSerLysAsnValIleLeuLysGluPheMetArgPheLysValArgMetGluGly 20
            ||:||:::::||||||||||||||||||||||||||||||:::::||
Db   1 ATGGCCTCCTCCGGAGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy   21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
            :|||||||||||||||||||||||||||||||||||||||||||
Db   61 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy   41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            |||
```

Left column

```
Db   121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGGCGGCCCCCTTGCCCTTGCCCTGGGGACATC 180

Qy       LeuSerProGlnPheGlnTyrGlySerLysValTyrVallysHisProAlaAspIlePro 80
Qy    61 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGGCTACGTGAAGCACCCCGCAGATCCCC 240
Db   181 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy       AspTyrLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy    81 GACTACTTGAAGGCTGTCCTTCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300
Db   241 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy       AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   101 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy       LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   121 AAGGTGAAGCTGCGCGGCGTCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy       MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   141 ATGGGCTGGGAGGCCTCCACCGAGCGCGATTACCCCGAGGACGGCGTGCTGAAGGGCGAG 480

Qy       IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   161 ATCCACAAGGCTCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC 540

Qy       TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   181 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy       IleThrSerHisAsnGluAspTyrThrIleValGluLeuValArgThrGluGlyArg 220
Db   201 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACTCGTGAGGACAGAGGGCCGC 660

Db   221 His 221
Db   661 CAC 663
```

RESULT 14
US-11-218-880-9
; Sequence 9, Application US/11218880
; Publication No. US20060003420A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/11/218,880
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 9
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: "mRP1"
US-11-218-880-9

Alignment Scores:
Pred. No.:
Score: 1.36e-110
Percent Similarity: 1021.00
Best Local Similarity: 89.14%
Query Match: 86.88%
                          84.10%

Length: 678
Matches: 192
Conservative: 5
Mismatches: 24
Indels: 0

Right column

```
DB:                      7          Gaps:       0

US-10-006-922A-12 (1-225) x US-11-218-880-9 (1-678)

Qy       MetArgSerSerLysAsnValIleIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db     1 ATGGGCCGTCCTGGAAGGTCATCATCAAGAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy       ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db    21 TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy       HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db    41 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGGGCGGCCCCTTGCCCTTCGCCCTGGGACATC 180

Qy       LeuSerProGlnPheGlnTyrGlySerLysValTyrVallysHisProAlaAspIlePro 80
Db    61 CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGGCTACGTGAAGCACCCCGCAGATCCCC 240

Qy       AspTyrLysGlyLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db    81 GACTACTTGAAGGCTGTCCTTCCCGAGGGCTTCAAGTGGGAGAGGGTGATGAACTTCGAG 300

Qy       AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db   101 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 420

Qy       LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db   121 AAGGTGAAGCTGCGCGGCGTCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy       MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db   141 ATGGGCTGGGAGGCCTCCACCGAGCGCGATTACCCCGAGGACGGCGTGCTGAAGGGCGAG 480

Qy       IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db   161 ATCCACAAGGCTCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGCATC 540

Qy       TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db   181 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACAGCAAGCTGGAC 600

Qy       IleThrSerHisAsnGluAspTyrThrIleValGluLeuValArgThrGluGlyArg 220
Db   201 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACTCGTGAGGACAGAGGGCCGC 660

Qy   221 His 221
Db   661 CAC 663
```

RESULT 15
US-10-209-208-80
; Sequence 80, Application US/10209208
; Publication No. US20050244921A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; APPLICANT: Geoffrey Baird
; TITLE OF INVENTION: FLUORESCENT PROTEIN VARIANTS AND METHODS
; TITLE OF INVENTION: FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP2
; CURRENT APPLICATION NUMBER: US/10/209,208
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSEQ for Windows Version 4.0

```
;  SEQ ID NO 80
;    LENGTH: 678
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Polynucleotide encoding DsRed polypeptide variant
;    OTHER INFORMATION: "mRFP1.1"
US-10-209-208-80

Alignment Scores:
Pred. No.:
Score:                  1.58e-109        Length:           678
Score:                  1012.00          Matches:          190
Percent Similarity:     88.69%           Conservative:     6
Best Local Similarity:  85.97%           Mismatches:       25
Query Match:            83.36%           Indels:           0
DB:                     6                Gaps:             0

US-10-006-922A-12 (1-225) x US-10-209-208-80 (1-678)
```

```
Qy      1 MetArgSerSerLysAsnValIleLeuGluProPheMetArgPheLysValArgMetGluGly 20
            | || ||  ||:::::|||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGCTTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCCGCATGGAGGGC 60

Qy     21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyGly 40
            :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCGTGAACGGGCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC 120

Qy     41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATC 180

Qy     61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC 240

Qy     81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
            ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db    241 GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG 300

Qy    101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360

Qy    121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC 420

Qy    141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
            |||||||||||:::||||||||||:|||||||||||||||||||||||||||||||||||
Db    421 ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGAGGACGGCGTGCTGAAGGGCGAG 480

Qy    161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
            ||||||||||||||||||||:||||||||||||||||||||||||||||||||||:::|||
Db    481 ATCAAGATGAGGCTGAAGGACGGGCCCCACTACCTGACGGCCGAGGTCAAGACCACC 540

Qy    181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTAGACAGCAAGCTGGAC 600

Qy    201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
            ||||||||||:::|||||||||||||||||||||||||||:|||||||||||| |||||
Db    601 ATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGC 660

Qy    221 His 221
            |||
Db    661 CAC 663
```

Search completed: January 12, 2006, 15:22:24
Job time : 388 secs